Copyright

by

Sooin Jang

2015

The Dissertation Committee for Sooin Jang Certifies that this is the approved version of the following dissertation:

Mechanism of repair of Mu DNA insertions

Committee:

Rasika M. Harshey, Supervisor

David L. Herrin

Makkuni Jayaram

Kyle M. Miller

Tanya T. Paull

Mechanism of repair of Mu DNA insertions

by

Sooin Jang, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2015

Dedication

To my parents

Acknowledgements

I deeply appreciate my supervisor, Dr. Rasika M. Harshey. She has not only given me this fascinating project, but also constantly supported and encouraged me throughout my graduate years. I have truly enjoyed my work, and been inspired and motivated to be an independent researcher under her great mentorship. This work could never be accomplished without her patient guidance. I would also like to express my sincere gratitude to the members of my dissertation committee, Dr. Makkuni Jayaram, Dr. Tanya T. Paull, Dr. Kyle M. Miller and Dr. David L. Herrin for their invaluable advice and constructive feedback on my work. I also thank Dr. Steven J. Sandler, Dr. James R. Walker and Dr. Alexandra L. Blinkova for their priceless strains and great advice that have made this project much better.

I am grateful to all my past and present colleagues for their friendship and helps, especially to Dr. Wonyoung Choi and Dr. Rudra Saha for their insightful advice and guidance over my graduate years. I also appreciate the Jayaram lab members for sharing their equipment and materials without hesitation for my experiment, especially thanks to Chien Ma for her great experimental advice during my graduate work.

Most of all, I deeply appreciate my parents for their endless love and sacrifice over my life. Their unconditional support and confidence in me have enabled me to pursue my dreams. Lastly, I thank to my dearest sister and brother for their love and support.

Mechanism of repair of Mu DNA insertions

Sooin Jang, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Rasika M. Harshey

Transposable elements are ubiquitous, occupying as much as 85% of the genome of some species, and nearly 50% of the human genome, and causing DNA disruptions, mutations, and rearrangements. While transposable elements move in a variety of ways, their transposases cut and join DNA in a similar manner, all ultimately creating short flanking gaps in the target DNA. Transposition is not complete until these gaps are filled, yet this last step is still a black box. Using transposable phage Mu, we have made new discoveries that shed light on this step. We find that Mu recruits the Pol III replisome, and not gap-filling polymerases, for gap repair. Taking advantage of the high efficiency of Mu transposition and of a unique feature of its transposition intermediate, we made the surprising observation that the transpososome waits for the replisome to begin repair. When a fork runs into a gap, a double strand break (DSB) is expected: we demonstrated fork-dependent DSBs proximal to Mu. This result is consistent with genetic studies showing that recovery of Mu insertions requires the DSB repair pathway. These findings immediately suggest a model wherein the double strand break is exploited by the

transpososome for coordinated repair of the two flanking gaps by the two Pol III subunits, without replicating the intervening transposon DNA. Such a maneuver constitutes a novel DNA transaction for the polymerase and for repair. These findings are of broad significance because Mu, retrotransposons and retroviruses (e.g. HIV-1) transpose by a similar mechanism.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

GENOME STABILITY AND PLASTICITY

Genomes of all living cells are stable and well-organized. Stable genomes are maintained by the action of numerous structural and functional proteins that ensure replication fidelity, faithful chromosome segregation and repair of mutations. Mutations are caused by replication errors and by DNA damage (Papamichos-Chronakis and Peterson, 2013).

DNA damage stems from both endogenous and exogenous sources. Endogenous damage can be the result of metabolic or hydrolytic processes that release compounds such as reactive oxygen species and alkylating agents. Exogenous damage can arise by ultra-violet light as well as chemicals. Cells have many mechanisms to detect and repair such damage, so that the genome is protected and stably transmitted (Sinha and Hader, 2002; Barnes and Lindahl, 2004; Poirier, 2004; Jackson and Bartek, 2009).

While organisms need a high degree of genetic stability to maintain their species identity over long evolutionary times, they must also retain some mutability in order to acquire new traits to adapt to changing environments (Kasuga and Gijzen, 2013; Darmon and Leach, 2014). Processes contributing to mutability include DNA repair mechanisms that are not 100% efficient, for example, proofreading of replication errors (Drake et al., 1998). Polymerase slippage at some DNA sites results in addition or deletion of bases that also go unrepaired (Viguera et al., 2001). Duplication or amplification of genes occasionally arises as a product of misaligned chromosomal recombination or activation

of particular genes or sequences (Zhang, 2003; Santarius et al., 2010). When DNA damage is unrepaired, DNA replication stalls at the damage sites, giving rise to double-strand breaks (DSBs), which can introduce the synthesis of error-prone polymerases that generate new mutations (Goodman, 2002; Ponder et al., 2005). Other drivers of genome diversity include the evolution of adaptive immune systems. In bacteria and archaea, these are the CRISPR systems (clustered regularly interspaced short palindromic repeats), where invading viral DNA is captured into CRISPR loci to confer immunity to future infections, at the same time increasing diversity of the host genome as CRISPRs evolve during repeated viral infections (Horvath and Barrangou, 2010). Similarly, in vertebrates, immunoglobulin gene rearrangements by the V(D)J recombination system confers genetic diversity while providing adaptive immunity against invading pathogens (Hiom et al., 1998). Besides these diverse mechanisms for introducing genetic variation, a major source of genetic mutations and rearrangements is the movement of transposable elements, the subject of this study. They will be introduced in detail in the next section.

TRANSPOSABLE ELEMENTS

Overview

Transposable elements (TEs, or transposons) are discrete segments of DNA that can move from one location in the genome to another. For this reason, they have also been called 'jumping genes' or 'mobile DNA'. The main distinction between TEs and the other kinds of mobile DNA (e.g. λ-like family, P1, self-splicing introns) is that they can insert into non-homologous regions within a chromosome, whereas the other elements can only move into regions of homology (e.g. *att* site of λ integration or *lox* site of P1 recombination) (Craig, 2002). The ability of TEs to insert at virtually any site is responsible for their widespread dissemination in nature.

TEs were first discovered in the maize chromosome by Barbara McClintock in the 1940s (McClintock, 1950). Her innovative findings were initially dismissed by the scientific community as a special case perhaps unique to maize. Acceptance of her ideas began when TEs were spotted in phage and in bacteria during the 1970s and 1980s, and their presence was confirmed by emerging advanced techniques of recombinant DNA such as Southern blotting and polymerase chain reaction (PCR) (Fedoroff, 1994). Their stunning numbers in the genome came to light with NextGen sequencing methodology, which allowed rapid sequencing of whole genomes. These studies revealed the presence of TEs practically everywhere, from unicellular to higher organisms, where they comprised various portions of the genome – 1-2% in *E. coli*, 3% in yeast, 15% in *Drosophila*, 50-80% in plants, and 45% in humans (Kidwell and Lisch, 2000; Lander et al., 2001; Schnable et al., 2009). It was clear that TEs were restructuring all genomes.

The start of the 21st century has seen the construction of so many massive genome databases – that the neologism "-omics" has been adopted to describe them. Advances in

DNA sequencing led to Genomics, while advances in mass spectrometry and RNA-sequencing led to Proteomics and Transcriptomics, respectively. Together, these advances have reinforced the profound impact of TEs on genome structure, gene regulation (or epigenesis), and ongoing genome evolution (Fedoroff, 2012). Remarkably, recent studies have found that the presence and movement of TEs affects neuronal development. On the one hand, movement of TEs in the *Drosophila* brain has been shown to drive neuronal heterogeneity (Perrat et al., 2013). On the other hand, TEs are highly activated in a *Drosophila* brain during normal aging, implying influence of TEs on age-dependent loss of neuronal function (Li et al., 2013). In human embryonic stem cells, the movement of TEs alters the transcriptional level of many critical regulatory genes, suggesting that TEs play a role in rewiring core regulatory networks during cell differentiation (Kunarso et al., 2010). Currently, many studies with various human organs are examining the relationship of TEs to specific tumorigenesis pathways (Chenais, 2013; Takeda et al., 2015). Without a doubt, Mobilomics – mapping all the TEs in the genome and analyzing their dynamic behavior – is a new resource, and is expected to increase our understanding of the evolution of complex traits.

Despite the widespread documentation of the impact of TEs on genome structure and function, much less is known about how their movement is regulated or how they interact with, and exploit host pathways for their movement.

Classification

TEs can be categorized into many different types on the basis of their structure and mechanism of movement. At present, they are grouped into two major classes according to whether their transposition requires an RNA (Type I) or a DNA (Type II)

intermediate (Fig. 1.1). Type I elements (retrotransposons, or RNA transposons) are mostly found in mammals, and copy themselves in two stages – 1) transcription from DNA to RNA, 2) reverse transcription from RNA back to DNA. The DNA copy is then inserted into a new site on the genome. In some elements, the RNA intermediates themselves are active and insert directly into a target without reverse transcription (e.g. Group II introns) (Prak and Kazazian, 2000; Slotkin and Martienssen, 2007).

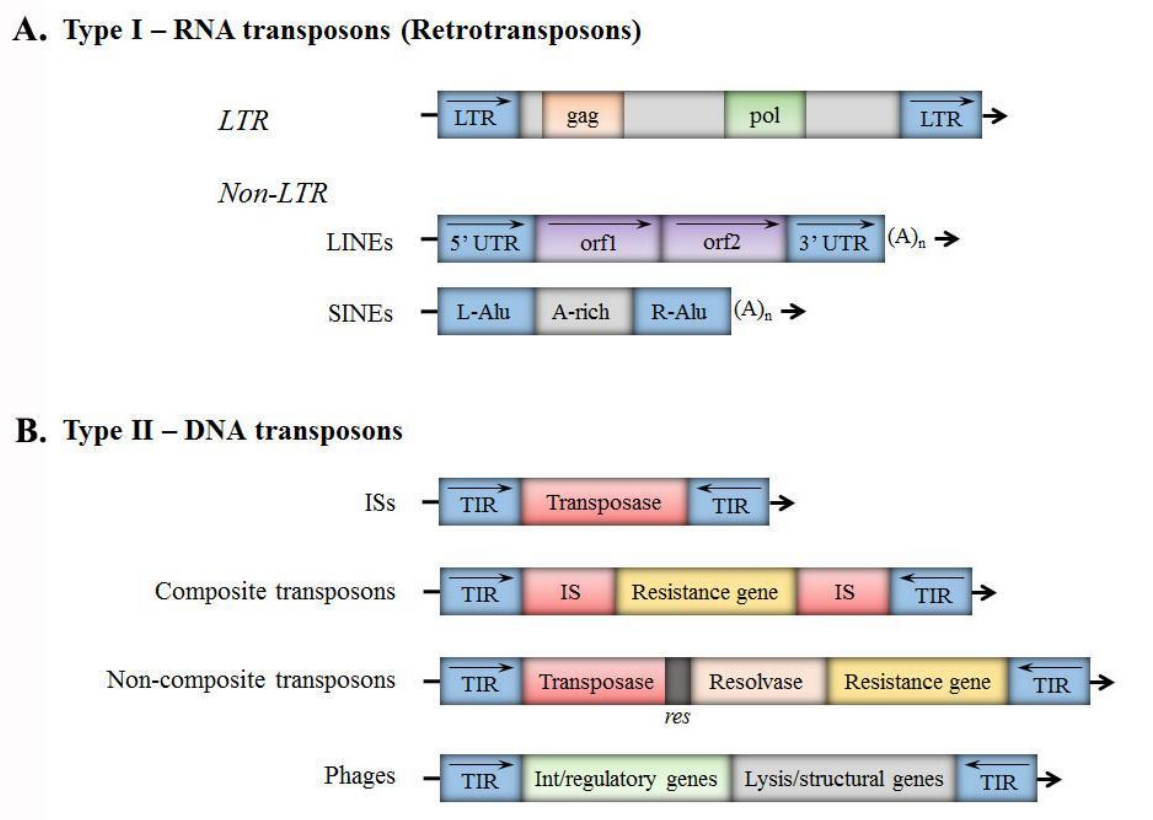


Figure 1.1 Classification of transposable elements (TEs).

See text for details.

Type I elements can be divided into two subtypes, based on the presence or absence of long-terminal repeats (LTR) (Fig 1.1A). LTR retrotransposons encode a reverse transcriptase that transcribes the RNA to DNA. Retroviruses (e.g. HIV-1) belong to the LTR category in that they also have an LTR region at their ends and encode reverse transcriptase. Upon infection, retroviruses convert their RNA genomes into DNA by help of reverse transcriptase, and integrate the DNA copy into the host genome. The integrated DNA (or provirus), when activated, reproduces by transcribing the DNA into RNA copies which correspond to the original retroviral genome. This life cycle of retroviruses is very similar to that of prokaryotic transposable phages which can remain as a provirus until activated. The Non-LTR retrotransposons are further divided into LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements). LINEs are ~ 4-7 kb in length, are transcribed by RNA polymerase II, and encode the enzymatic activities required for their mobility. SINEs, non-coding sequences of ~ 80-400 bp, are transcribed by RNA polymerase III, and require LINE-encoded endonuclease and reverse transcriptase for their movement (Dewannieux et al., 2003).

Members of the other class of elements (Type II, DNA transposons) are found in both prokaryotes and eukaryotes, and do not need an RNA intermediate to transpose. They can be also divided into several subtypes; ISs (Insertion sequences), composite transposons (e.g. Tn10), non-composite transposons (e.g. Tn3, Tn7), and phages (e.g. Mu) (Fig 1.1B). ISs are the simplest of all transposons (750 bp-1600 bp); they harbor only one or two genes that encode the transposases required for movement. IS termini are usually 10-40 bp in length with identical or almost identical indirect repeats. Composite transposons have a pair of direct or inverted IS elements flanking genes conferring some advantage to the host such as antibiotic or toxin resistance. For instance, Tn10 is a 9.3 kb long transposon, which encodes tetracycline resistance and is flanked by a pair of

inverted IS10 elements at its ends (Mahillon and Chandler, 1998). Non-composite transposons are large elements (more than 5 kb), lack flanking IS elements, and have a pair of inverted repeats of 38-40 bp at their ends. They encode a transposase (TnpA) and a site-specific recombinase of the resolvase (TnpR) or integrase (TnpI) family (Craig, 2002). Non-composite transposons also carry antibiotics resistance or other beneficial genes and contain a pair of terminal inverted repeats (TIR) instead of IS elements.

Another subtype of DNA transposons is phages like Mu and its relative D108. These bacterial viruses use transposition for generating prophages as well as for amplifying their genome. Their complex genomes lack true TIRs, although their ends are specifically recognized by the transposase to catalyze transposition. They also contain *cis*-acting regulatory elements such as a transposition enhancer to enhance amplification of their genomes and a gyrase binding site thought to nucleate supercoiling and promote pairing of the ends, which also enhances transposition efficiency. In addition, these elements code for proteins required for assembling phage particles and for lysis of the host to release these particles (Chaconas and Harshey, 2002).

Mechanism of transposition

Transposons move in different ways depending on their donor/target structure and the catalytic activity of their transposase. Their movement can be broadly categorized as either replicative or non-replicative transposition depending on whether or not their DNA is duplicated during transposition (Fig 1.2). In replicative transposition, one copy of the transposon remains at its original site while the other copy inserts at the new site (Fig 1.2A). Thus transposition is accompanied by an increase in the number of copies of the transposon. On the other hand, non-replicative transposition occurs by movement of the

original DNA from one place to the other without replication (Fig 1.2B). Nearly all TEs, including bacterial transposons, P-elements of *Drosophila*, Ac/Ds of maize, as well as eukaryotic DNA Tc*1/mariner* family and retrotransposons, belong to the non-replicative category (Haren et al., 1999).

Replicative transposition is performed by a limited number of transposons. These include the Tn3 transposon family and transposable phages (Ohtsubo et al., 1981; Chaconas and Harshey, 2002). The latter use both replicative and non-replicative mechanisms, depending on their life cycle. For instance, transposable phages Mu and D108 insert their DNA into the host chromosome by non-replicative transposition during lysogenic growth, and amplify their genomes by replicative transposition during lytic growth (Symonds et al., 1987).

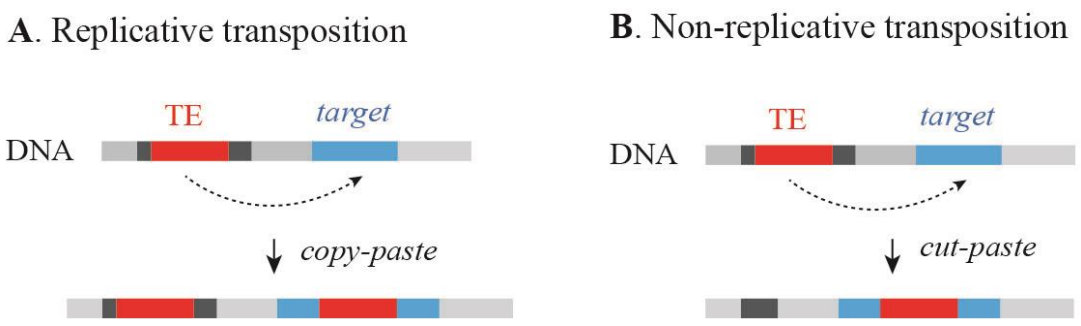


Figure 1.2 Two modes of transposition.

(**A**) TE (red) inserted in donor DNA (dark grey) is duplicated into new target DNA (blue), also referred to as "copy and paste" mechanism. TEs amplify their copies within a host genome by this mechanism. (**B**) TE is cut out of its original location and inserted into a new target location, also referred to as "cut and paste" mechanism.

Variety of transposases

There are five major transposase families, each with a distinct catalytic mechanism. These are called DDE, Tyrosine (Y), Serine (S), Rolling-circle (RC), and Reverse transcriptase/endonucleases (RT/En) transposase families (Curcio and Derbyshire, 2003). Among them, DDE-transposases are the most prevalent. DDE stands for a conserved triad of amino acids – Asp (D), Asp (D), and Glu (E) – which coordinate divalent metal ions necessary for the DNA cleavage and joining (strand transfer). Figure 1.3 shows different patterns of transposition catalyzed by DDE-transposases. The Tn7 transposase cleaves both DNA strands at their ends, excising Tn7 from the donor. The excised Tn7 inserts into new target site, generating a strand transfer (ST) intermediate that is subsequently repaired (Craig, 1996). Similarly, Tn10 also makes double strand cleavages at both ends, but goes through a hairpin intermediate prior to ST (Kennedy et al., 1998). Several TEs follow this mechanism (e.g. IS4, IS10) (Mahillon and Chandler, 1998). Mu and Tn3 cleave only single DNA strands at their ends and insert into target site to form a branched ST intermediate linked to both donor and target DNA (Grindley, 1983; Lavoie and Chaconas, 1996). Other TEs undergo alternative routes to form ST intermediates such as circularization of the excised elements (e.g. IS3) (Mahillon and Chandler, 1998) or transcription/reverse transcription before insertion (e.g. retroviruses) (Coffin, 1992). After ST, the intermediate either trims off the short donor sequence still attached to each 5'end, completing transposition by filling in the short gaps generated in the target on either side of the TE (Tn7, IS3, retroviruses), or undergoes target-primed replication that resolves the branched molecule, filling the flanking gaps and duplicating the transposon (replicative transposition; Mu, Tn3) (Haren et al., 1999). Both mechanisms are thought to rely on host proteins to complete the transposition. As

described above, Mu is unique in using both replicative and non-replicative transposition mechanisms, which will be described in detail in the section on Phage Mu.

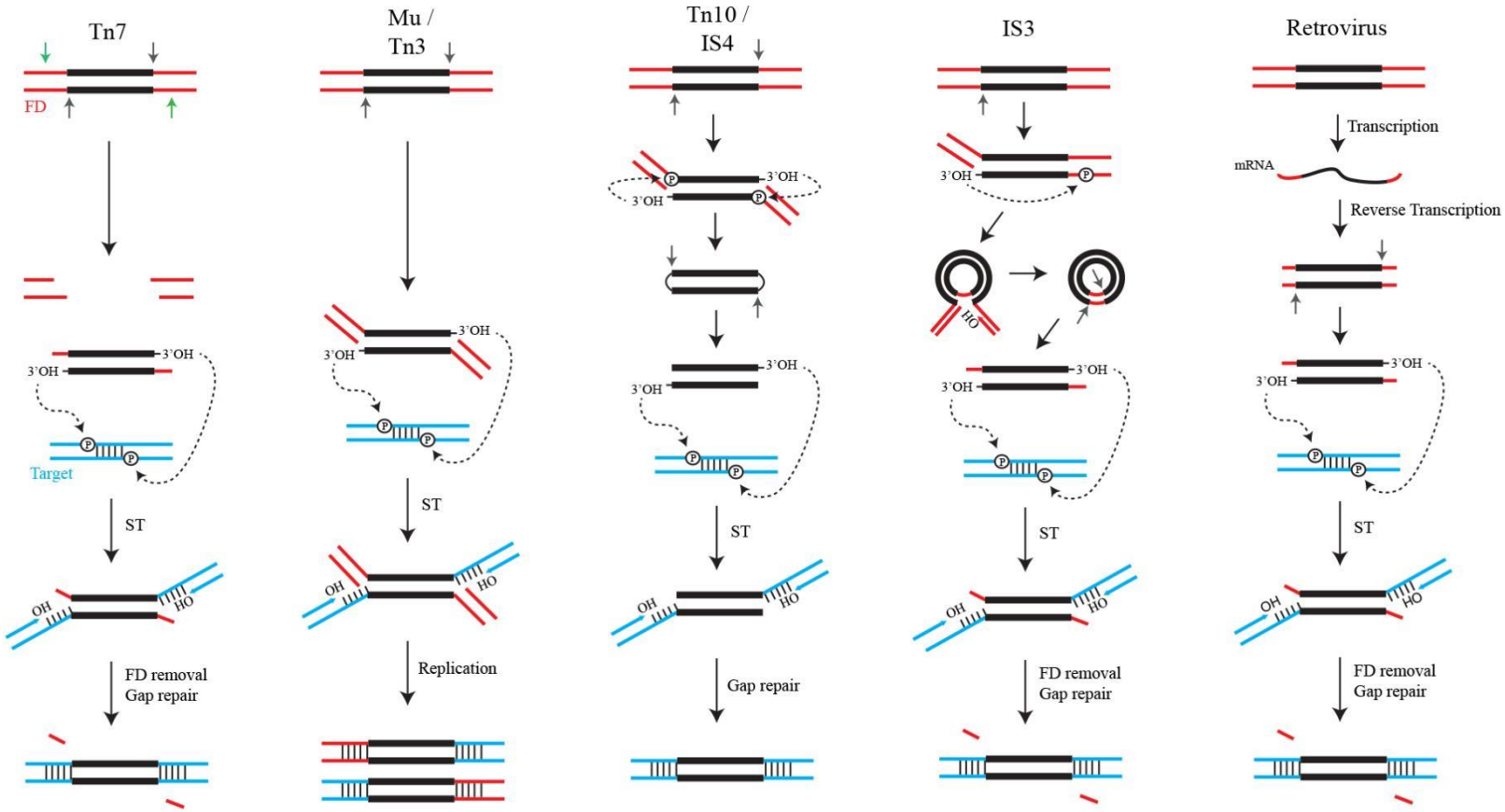


Figure 1.3 Diverse mechanisms of DNA transposition by DDE transposases.

TE sequences are represented by black lines flanked by red donor DNA (FD), and target sequences are cyan. Initial water-mediated cleavages on the donor DNA are indicated by short arrows; these generate 3'OHs. The 5' endonucleolytic cleavages of non-transferred strands are a different reaction indicated by green arrows. Nucleophilic attack by the 3'OHs on target phosphodiester bonds (P) is indicated by dotted lines with long arrows. This reaction, which joins the donor and target, is called strand transfer (ST). Vertical comb lines (black) shown on target DNA represents short stretches of nucleotides which will be converted into target duplications after ST. This figure was modified from Haren et al., 1999 and Hickman et al., 2010.

Unity of strand transfer

Despite the apparent variety of transposition mechanisms, the phosphor-transfer chemistry of cleavage and strand transfer is conserved. All transposases commonly catalyze two major reactions – 1) strand cleavage of the 3' terminal phosphodiester bonds in the element to expose free 3'OH groups, using water as a nucleophile, 2) strand transfer, where the TE DNA 3'OHs generated in the cleavage step act as nucleophiles to attack phosphodiester bonds in the target (transesterification) (Engelman et al., 1991; Mizuuchi, 1992). Both steps are catalyzed within the DDE active site, where these residues position divalent metal ions (Mg^{2+} or Mn^{2+}) for successive activation of the nucleophile – first water and then the DNA hydroxyl. Both steps are coordinated by pairing of the TE ends within the transpososome. Coordinated ST of the cleaved TE ends into the two strands of target DNA creates 5' staggered cleavages in the target separated by 2-9 nucleotides (depending on the transposase), joining the two DNAs without requiring an external source of energy (Mahillon and Chandler, 1998). The ST intermediate is resolved by cellular DNA replication or repair machineries.

The chemistry of transposition was first elucidated using an *in vitro* system established for Mu (Mizuuchi, 1983). Despite little or no similarity in their primary sequence with the Mu transposase, many other transposases/integrases have structurally similar active sites, and share a common catalytic mechanism of phosphoryl transfer. For instance, crystal structures of retroviral integrases (HIV, ASLV) (Bujacz et al., 1996; Chen et al., 2000) and RAG1-RAG2 of V(D)J recombination (Kim et al., 2015) reveal that the conserved DDE motif coordinates a metal ion similar to that seen in the Mu transposase (Rice and Baker, 2001). Also, like the Mu transposase (Aldaz et al., 1996; Savilahti and Mizuuchi, 1996; Namgoong and Harshey, 1998), other transposases (e.g.

Tn5, Tn7, IS630 family) and retroviruses (e.g. PFV (prototype foamy virus)) show a topologically similar *trans* arrangements of catalytic subunits i.e. the subunit bound to one end catalyzes reaction chemistry at the other end (Gueguen et al., 2005; Hare et al., 2010). In the Mu transpososome, the target DNA is accommodated at the active sites by being sharply bent in order to allow the cleaved ends to attack two closely spaced target phosphodiester bonds (Montano et al., 2012). The bent target conformation seen in the Mu complex is also observed in other transposons, for example in Tn10 and in the retrovirus PFV (Pribil and Haniford, 2003; Maertens et al., 2010). The structure of the ST complex has been exploited in the development of potential inhibitors of the viral integration (Hare et al., 2010). One such inhibitor, Raltegravir, is currently marketed by Merck as an anti-HIV drug (Summa et al., 2008).

Post-strand transfer

The cleavage and joining reactions of transposition generate an ST intermediate with short single-stranded gaps on the target DNA due to the staggered cleavage of the target (Fig 1.3). There are two ways this intermediate can be resolved/repaired. Mu and Tn3 employ target primed replication using the flanking target 3'OHs as primers; this duplicates the TE (replicative transposition). A second mechanism is localized replication confined to the gaps i.e. gap repair (non-replicative transposition). This function could be performed by gap-filling polymerases, but whether this is indeed so is not yet established.

Hypothetically, the post-ST events in non-replicative transposition can be divided into three major steps – 1) trimming the flanking DNA (FD) overhangs at the 5' ends, 2) filling the single-stranded gaps, 3) ligation of the remaining nicks (Smith and Daniel, 2006).

FD trimming: The length of FD is variable in different TEs (from few nucleotides about 2-3 bp to over 1 kb), so FD trimming may exploit endonuclease or exonuclease activities. Several candidates for these activities have been proposed. Since repair pathways such as mismatch repair (MMR), nucleotide excision repair (NER), or base excision repair (BER) similarly requires a nuclease, a polymerase and a ligase, transposons might likely exploit these repair proteins (Yoder and Bushman, 2000; Daniel et al., 2003; Espeseth et al., 2011; Yoder et al., 2011).

Filling single-stranded gaps: The gap-filling polymerase associated with the completion of Okazaki fragments on the lagging strand is a prime candidate for filling transposon gaps. In *E. coli*, this polymerase is Pol I (or PolA). In support of this possibility, Tn5 and Tn10 showed a low frequency of transposition in *polA*-deficient strains (Clements and Syvanen, 1981; Sasakawa et al., 1981). However, the transposons used in these studies were carried on λ and the results could be related to the effect of *polA* on λ replication. Also, the *polA* mutants used (e.g. *polA1*, *polA11*, *polAex2ts*, *polA34ts*) were all point mutants, with residual activity of either the exonuclease or polymerase (Sasakawa et al., 1981). This was necessitated because a complete deletion reduces cell viability in rich media (Joyce and Grindley, 1984). Because of this, null mutant of Pol I were not tested. Recovery of Mu insertions in a *polA1* mutant were reported to be reduced 3-8 fold, implying that Pol I is not essential for Mu DNA repair, and that other repair proteins might also be involved (McBeth and Taylor, 1983).

In mammals, two members of DNA polymerases, Pol λ and μ, have been considered as candidates for gap repair of transposition events because both are involved in filling single-stranded gaps during non-homologous end joining (NHEJ), one of two major pathways of DNA double-strand break (DSB) repair (Li et al., 2001; Skalka and Katz, 2005). Other reports suggest that polymerase β or δ, which is used primarily for

BER or for lagging strand synthesis during replication, is required for the gap repair of transposition *in vitro* (Yoder and Bushman, 2000; Van Cor-Hosmer et al., 2013). Support for these ideas is lacking *in vivo*.

The focus of this dissertation is to identify host factors involved in the post-ST repair of Mu insertions during non-replicative transposition, and in so doing, to fill the gap in our knowledge of this essential step required to complete transposition.

PHAGE MU

Overview

Mu is a transposable phage, discovered by Larry Taylor in the early 1960s. Taylor christened this new phage Mu for 'mutator', because Mu lysogens were associated with host mutations (Taylor, 1963). The significance of Mu as a transposable element was revealed by pioneering studies by Ahmad Bukhari and Arianne Toussaint in 1970s (Harshey, 2012). The high efficiency of Mu transposition not only made it a powerful genetic tool for studying the DNA arrangements associated with the transposition, but also allowed the establishment of the first *in vitro* system for transposition (Mizuuchi, 1983). The *in vitro* studies were instrumental in advancing our knowledge of the chemical reactions of transposition and of our understanding of how all TEs move (Craig, 2002). The Mu *in vitro* system was also exploited for studying the mechanism of replicative transposition (Levchenko et al., 1995; Jones and Nakai, 1997; Nakai et al., 2001).

Life cycle

As a temperate phage, Mu can undergo either lysogenic or lytic growth depending on its life cycle (Fig. 1.4). During infection, the linear Mu virion genome is injected into the host along with a Mu protein called N, which binds to both ends of the flanking DNA (FD). N circularizes Mu DNA, holds the ends together non-covalently, and protects them against attack from host nucleases (Harshey and Bukhari, 1983; Puspurs et al., 1983; Gloor and Chaconas, 1986). Early transcription from the Mu genome produces the transposition proteins MuA and MuB (Wijffelman and van de Putte, 1974). MuA is the

transposase, which binds to both Mu ends and forms a multi-subunit transpososome (Chaconas and Harshey, 2002). MuB performs several roles, the most important one being capturing the target DNA and delivering it to the transpososome (Mizuuchi and Mizuuchi, 1993). After infection, the first integration event is special and distinct from all others that follow. During this initial event, Mu transposes by a non-replicative mechanism i.e. the FD is trimmed and the target gaps are repaired without replication of Mu (Liebart et al., 1982; Akroyd and Symonds, 1983; Chaconas et al., 1983; Harshey, 1984). The majority of infected cells undergo lytic growth. Of the survivors, ~1-10% are stable lysogens (Howe and Bade, 1975). During lytic growth, Mu DNA is amplified over 100-fold by replicative transposition (Chaconas et al., 1981). At the end of the lytic cycle, Mu copies are package into phage heads such that host DNA flanking the insertion is also packaged, as depicted in Figure 1.4 (Symonds et al., 1987).

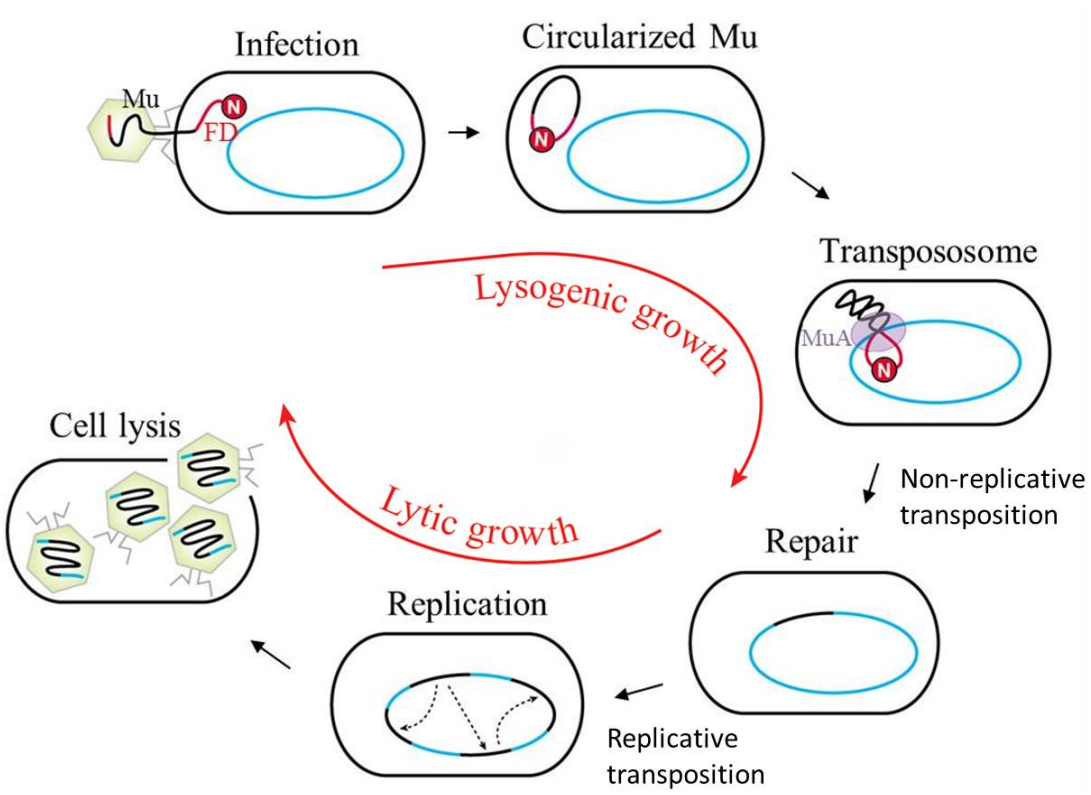


Figure 1.4 Life cycle of phage Mu.

Mu particle, yellow hexagon; Mu genome, black line; flanking DNA (FD), red line; MuN protein, red circle; MuA transposase, purple; host chromosome, blue circle. See text for details.

Mu genome structure

The DNA packaged in a Mu phage particle is about 39 kb long: Mu DNA is 37 kb, and the rest is flanking host or non-Mu DNA (Fig. 1.5). The size of non-Mu DNA is variable. At the left end, the FD can range from 60 to 150 bp, and at the right end from 1.5 to 3 kb. This variable feature is a property of the packaging machinery, which packages by a head-full mechanism (Bukhari and Taylor, 1975). After infection and integration, the FD attached to Mu is degraded during repair of the transposition event (Au et al., 2006; Choi and Harshey, 2010).

Organization of genes on the Mu genome is similar to that in other viruses. The early genes on the L end code for lysogenic regulation (*c*, *ner*) and transposition/replication (*A*, *B*). Additionally, ~5 kb of semi-essential (SE) genes including *kil* (killing the host) and *gam* (ortholog of eukaryotic Ku) are part of the early transcript (Morgan et al., 2002). The function of the majority of genes in this region is unknown. Phage structural genes for heads and tails are encoded in rest of the genome (Symonds et al., 1987). Besides the L and R ends, an essential *cis* element for transposition is the enhancer (E) located between *c* and *ner*. The MuA transposase binds to all three sites for assembly of the catalytically active transpososome (Chaconas and Harshey, 2002).

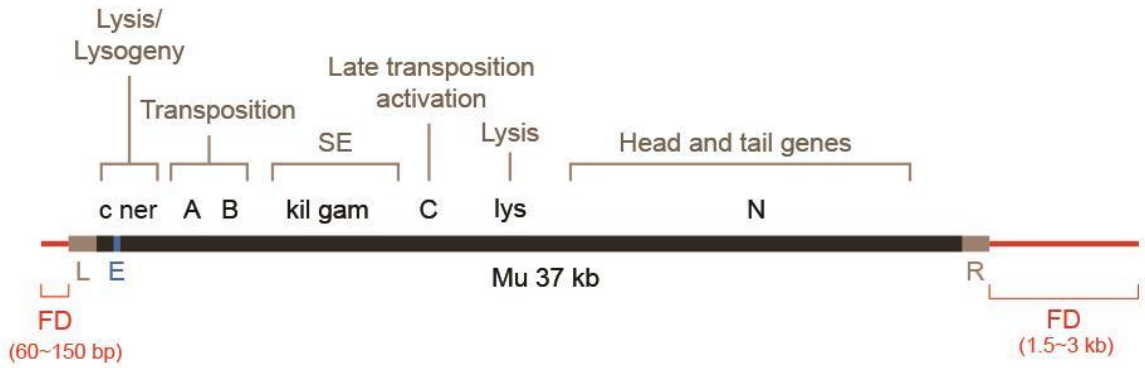


Figure 1.5 The structure of the packaged Mu genome.

Mu DNA (black), is flanked on the left and right (L and R, grey) by packaged host DNA (FD, red). The FD length is variable, shorter on the L end, longer on the R end. The enhancer (E, blue), is located between the lysogenic repressor gene *c* and lytic regulator gene *ner*. A few key genes and their function are indicated on top of the Mu genome.

Transposition mechanism

Most of our knowledge of Mu transposition is derived from *in vitro* experiments using mini-Mu plasmids as donors, other plasmids as target, and addition of Mu and *E. coli* proteins necessary for transposition (Mizuuchi, 1983). Our knowledge of post-ST events have come from monitoring resolution of the ST intermediate with added replication proteins, nucleases, and cell extracts (Kruklitis and Nakai, 1994; Jones and Nakai, 1997; Choi et al., 2014a). As shown in Figure 1.3 for Mu, transposase MuA mediates single-strand cleavages at Mu ends followed by strand transfer of the cleaved ends into target DNA (Mizuuchi, 1992); the latter reaction is greatly assisted by MuB protein (Fig. 1.6) (Mizuuchi and Mizuuchi, 2001). The resulting branched ST joint is resolved by two different mechanisms depending on the phase of the Mu life cycle. During lytic growth, where the Mu genome is amplified, the ST intermediate is resolved by target-primed replication. *In vitro* experiments have revealed a highly choreographed series of steps during which the transpososome is disassembled by ClpX and exchanged with series of known (IF2-2) and unknown proteins in preparation for assembly of the Restart replication complex or PriA primosome, which loads the Pol III holoenzyme at one Mu end, replicating across Mu to resolve the ST product (Kruklitis et al., 1996; Nakai et al., 2001; North and Nakai, 2005; North et al., 2007; Abdelhakim et al., 2008). During integration of infecting Mu, the ST intermediate is not resolved by target-primed replication. Instead, the FD is degraded and the flanking target gaps are somehow repaired. The bulk of the FD degradation is carried out by RecBCD both *in vivo* and *in vitro* (Fig. 1.6) (Choi and Harshey, 2010; Choi et al., 2014a). *In vivo,* FD degradation is dependent on a cryptic endonuclease activity (MuA_{Nuc}) harbored within the C-terminal domain of MuA (Wu and Chaconas, 1995), as well as on ClpX (Choi and Harshey,

2010). These requirements could be bypassed *in vitro* (Choi et al., 2014a). The alternative choices for resolving the transposition intermediate, i.e. repair versus replication, must involve additional phage and host factors whose identity is not as yet established.

Thus far, only the MuA transposase, RecBCD, and ClpX have been identified as being essential for the post-ST repair of the non-replicative transposition event (Fig 1.6). However, there must be additional factors that dislodge the N protein protecting the Mu ends (Fig. 1.4), and fill the target gaps after the FD is shortened. Several uncharacterized orfs in the early (SE or semi-essential) regulatory region of Mu are potential phage candidates for assisting with 5' flap cleavage (Fig. 1.5), while the host gap-filling DNA polymerases are potential candidates for filling the 5 bp gaps flanking Mu insertions.

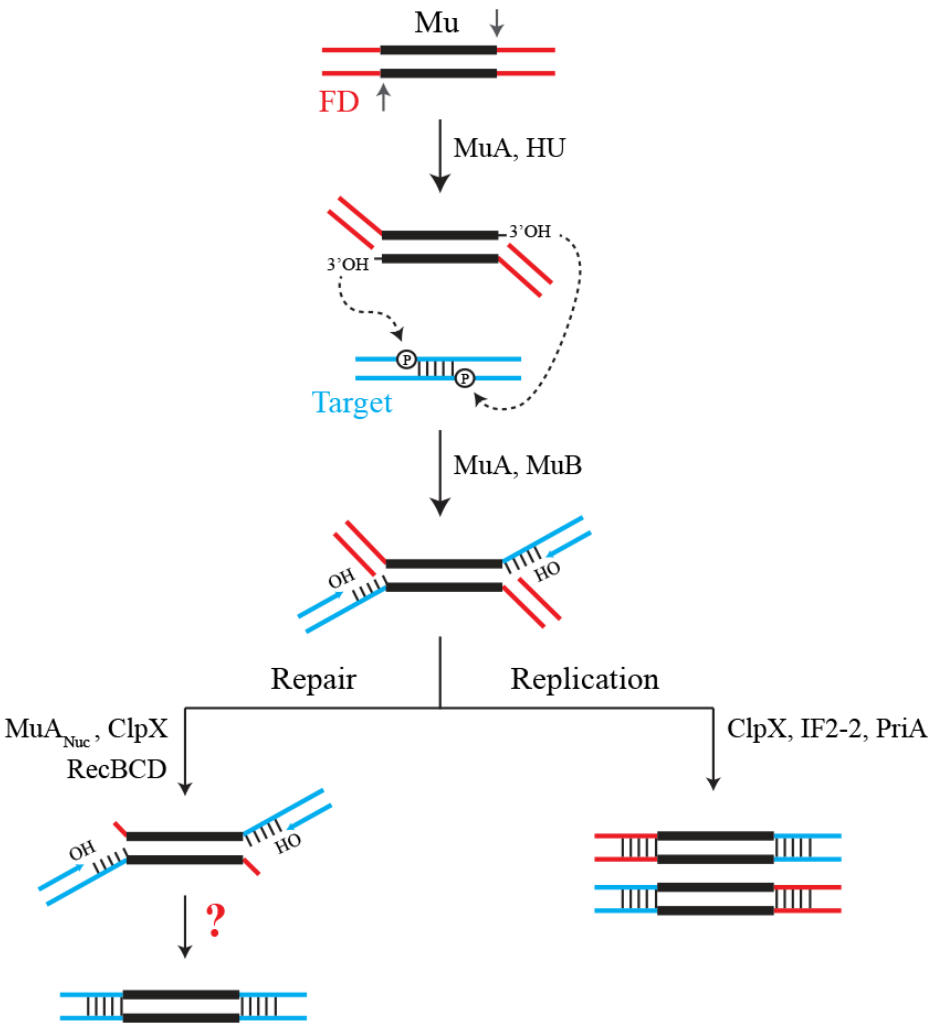


Figure 1.6 Known steps in replicative and non-replicative (repair) pathways of Mu transposition.

The transposase MuA, in the presence of *E. coli* protein HU, first introduces single-stranded cleavages at the ends. With assistance from MuB, the OHs at the cleaved ends are transferred by MuA to phosphodiester bonds spaced 5 bp apart in the target. The resultant branched strand transfer intermediate is processed alternately. During the lytic cycle, Mu transposition in intramolecular and occurs repeatedly within the circular *E. coli*

-Figure 1.6 *legend continued*-

chromosome. The ST intermediate is resolved by target-primed replication. ClpX, IF2-2 and other uncharacterized factors are required for disassembly of the transpososome followed by assembly of the PriA restart primosome on the Mu ends. During integration of infecting Mu, DNA transposition into the chromosome target is intermolecular. The branched strand transfer intermediate repaired by a two-step process, where RecBCD first degrades the bulk of the flanking DNA (FD), followed by shortening of the FD to +4 nucleotides. A cryptic endonuclease activity in MuA - MuA_{Nuc} - is implicated in an endonucleolytic event that generates the +4 product. ClpX is required for this reaction. It is not known how the remaining target gaps are filled.

SIGNIFICANCE AND GOAL OF THIS STUDY

The majority of known DNA and retrotransposons, including retroviruses like HIV-1, transpose by the non-replicative mechanism. A critical void in our knowledge of this pathway is how the target gaps are filled. Mu provides an excellent system to investigate this last step because of the high efficiency of non-replicative Mu transposition, where every infecting phage integrates into the *E. coli* chromosome. Infecting Mu also has an unusual feature – flanking DNA (FD) – which is degraded and repaired concomitant with gap repair. We have an *in vivo* assay to follow FD degradation (Au et al., 2006). Using this assay, the major goal of my work is to understand the gap repair step of non-replicative transposition.

Chapter 2. Materials and Methods

STRAINS

All strains used in this study are derivatives of *E. coli* K-12 and listed in Table 2.1. The Keio Collection (single-gene knockout library of 3,985 nonessential genes in *E. coli*) was obtained from the National BioResource Project, Japan (Baba et al., 2006; Yamamoto et al., 2009). The wild type strain in this collection is BW25113. They were propagated in LB media, except for *priA*, *dnaT* and *polA* mutants, which were grown in 56/2 minimal medium: 0.06 M Na_2HPO_4, 0.04 M KH_2PO_4, 0.02% $MgSO_4 \cdot 7H_2O$, 0.2% $(NH_4)_2SO_4$, 0.001% $Ca(NO_3)_2$, and 0.00005% $FeSO_4 \cdot 7H_2O$, 0.2% glucose, with casamino acids at 50 μg/ml (Willetts et al., 1969). P1 transduction was used to move mutations between strains (Miller, 1992).

E. coli Mu lysogen strains BU1717 or MH3491 were used to construct SJ17-SJ19 (Table 2.1), where a ~1 kb *cat* cassette was inserted downstream of the invertible G-segment on the Mu genome at nt 35,040, before *gin*, by the method of Datsenko and Wanner (Datsenko and Wanner, 2000). The SE deletion was similarly constructed; it removes nt 4,319-7,954 from the Mu genome, substituting the *cat* cassette in its place.

The Mu lysogen strain HM8305 was used to construct prophage Mu::TetO (SJ012) as follows. First, *kan* was introduced next to the TetO array in pRS306X112TetO at the SalI-SpeI sites (using primers P4/P5) to give pTetO(kan) (Table 2.2). pMuHF was then constructed to introduce ~500 bp homology corresponding to either side of the SE (Semi-essential region) (primers P6/P7 and P8/P9) in the prophage into which TetO*kan*

was to be substituted. Next, TetO*kan* was cloned within the two arms of the SE region in pMuHF to give pTetO(kan)-MuHF. This plasmid was digested with SalI and recombined into the Mu prophage to produce a deletion of the SE (~3.6 kb; 4319 bp to 7953 bp on the Mu genome) concomitant with a substitution (~6kb) with TetO*kan*, using λ-Red recombination (Datsenko and Wanner, 2000). The *kan* cassette was removed using Flp recombinase from pCP20. Finally, a non-essential region spanning ~2.4 kb at the right end of Mu (33,883 to 36,300 bp) was deleted by substitution with a *kan* cassette (primers P10 and P11), followed by removal of *kan* as before. The latter deletion was required to maintain a genome length that could be packaged into viable phage. All Mu phages used in this study carry the temperature-sensitive *ts*62 allele of the lysogenic repressor gene *c*.

Table 2.1 *E. coli* K-12 strains used in this study

Strain name	Relevant genotypes	Source (ref.)
Background strains and their derivatives		
MG1655	*rph1 ilvG rfb-50*	(Blattner et al., 1997)
SJ004	*dnaE486ts zae-502::Tn10*	J. Walker[a]
SMR14334	$P_{N25tetO}$ *gam-gfp*	(Shee et al., 2013)
SS1021	*dnaC28ts zji-202::Tn10*	(Withers and Bernander, 1998)
N4704	*rnhA::cat*	(Rudolph et al., 2013)
AU1054	*tnaA::Tn10 dnaA46*	(Rudolph et al., 2013)
AU1066	*tnaA::Tn10 dnaA46 rnhA::cat*	(Rudolph et al., 2013)
BU1384	F- Δ*pro lac* Su+	(Chaconas et al., 1984)
BU1382	*himA*Δ*82*	(Chaconas et al., 1984)
CW11	*ΔclpX::*kan	(Choi and Harshey, 2010)
SS996	*ΔattB::psulA-gfp*	(McCool et al., 2004b)
JC19328	Δ(*recA-srl*)306::Tn10	(McCool et al., 2004b)
SJ001	*ΔclpX::kan* from JW0428	This work
SS749	Δ(*recA-srl*)306::Tn10 *priA*2::kan	S. Sandler[b]
SS767	*malE*::Tn10 *lexA*3	(McCool and Sandler, 2001)
SS768	*priA*2::kan *lexA*3 *malE*::Tn10	(McCool and Sandler, 2001)

-Table 2.1 *continued*-

SS1411	zji-202::Tn10 *dnaT*822 Δ(*attB*)::p*sulA-gfp*	(McCool et al., 2004b)
SS1424	*dnaA46ts tnaA300::Tn10*	S. Sandler
SS1443	Δ(*priB*)302 Δ(*attB*)::p*sulA-gfp*	S. Sandler
SS1441	*priA*300 Δ(*attB*)::p*sulA-gfp*	(Boonsombat et al., 2006)
SS1448	*priA*2::kan Δ(*attB*)::p*sulA-gfp*	(McCool et al., 2004b)
SS2357	*Δ(polA)501::kan*	(Massoni and Sandler, 2013)
SS2400	*dnaC*809,820 p*sulA-gfp* thr+	S. Sandler
SS3085	*ygaD1::kan recAo1403 recA4136, 4155::gfp-901*(*recA-gfp*)	(Renzette et al., 2005)
SS3116	*priA*301 Δ(*attB*)::p*sulA-gfp*	S. Sandler
SS3403	*priC*303::kan Δ(*attB*)::*sulAp-gfp*	S. Sandler
SS4294	*malF::cam lexA3*	(McCool et al., 2004b)
SS4610	*lexA71::Tn5*	(McCool et al., 2004b)
SS6239	*dnaE486ts zae-502::Tn10*	(Massoni and Sandler, 2013)
SS6700	*dnaN159ts tnaA300::Tn10*	(Massoni and Sandler, 2013)
SS6698	*dnaX2016ts zbb-3055::Tn10*	(Massoni and Sandler, 2013)
SS6699	*dnaB8ts malE::Tn10-kan*	(Massoni and Sandler, 2013)
SS7086	zji-202::Tn10 *dnaC*809,820 *dnaT*822 Δ(*attB*)::p*sulA-gfp*	S. Sandler

-Table 2.1 *continued*-

SS7087	*priA*2::kan *dnaC*809,820 Δ(*attB*)::p*sulA-gfp*	S. Sandler
SS7346	*Δ(umuC)100::frt Δ(polB)100::frt Δ(dinB)100::frt*	S. Sandler
SS8775	Δ(*recBCD*)::kan	S. Sandler
SS8872	Δ(*recB*)100::kan	S. Sandler
BW25113	*rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	(Baba et al., 2006)
JW0428	*ΔclpX::kan*	(Baba et al., 2006)
JW2788	*ΔrecB::kan*	(Baba et al., 2006)
SJ005	*dnaE486ts zac1::Tn10* from SJ004	This work
SJ006	*ygaD1::kan recA-gfp* from SS3085	This work
HM8305	F' *pro lac*:Mu *cts62/Δpro lac his met rpsL* Mu^r	(Bukhari, 1975)
SJ009	Mu*ΔSE::tetO*$_{112}$*kan*	This work
SJ010	Mu*ΔSE::tetO*$_{112}$*frt* (*kan* removed)	This work
SJ011	Mu*ΔSE::tetO*$_{112}$*frt Δ*[*G-mom*]::*kan*	This work
SJ012 (Mu::TetO)	Mu*ΔSE::tetO*$_{112}$*frt Δ*[*G-mom*]::*frt* (*kan* removed)	This work
Other Mu prophages		
BU1091	F' *pro lac leu*::Mu *cts62* Amp	(Chaconas et al., 1984)

-Table 2.1 *continued*-

BU1717	F' *pro lac*::Mu *cts62 Bam*1066 Su-	(Chaconas et al., 1985)
CW45	MP1999 with *cat* at 35040 nt of Mu	(Choi and Harshey, 2010)
MH3491	Mu *cts62 Aam*1093 Su+	(O'Day et al., 1978)
MP1999	*recB recC sbcB malF*::Mu *cts62*	M. Pato[c]
SJ17	BU1717 with *cat* at 35040 nt of Mu	This study
SJ18	MH3491 with *cat* at 35040 nt of Mu	This study
SJ19	BU1717 with Δ*SE::cat* in Mu	This study

[a] University of Texas at Austin
[b] University of Massachusetts at Amherst
[c] University of Colorado health sciences center

Background parent strains are indicated in bold followed by their derivatives. SS996 is a derivative of JC13509. The genotype of JC13509 is *sulB103 lacMS286 ϕ80dIIlacBK1 argE3 hi-4 thi-1 xyl-5 mtl-1 rpsL31 tsx.* The *lacMS286 ϕ80dIIlacBK1* code for two partial non-overlapping deletions of the *lac* operon (Konrad, 1977; Zieg and Kushner, 1977).

PLASMIDS

Plasmids are listed in Table 2.2. The Gam-GFP fusion was amplified from the genome of strain SMR14334 using primers P1 and P2. The ~ 1.2 kb Gam-GFP product was digested by Sal I and Xba I and ligated into the same enzyme sites of the vector pRHA-113, where it was placed under the control of the rhamnose-inducible promoter. Similarly, a GFP-only control was amplified from same strain SMR14334 using primers P2 and P3. The resulting plasmids pGam-GFP and pGFP were verified by sequencing.

Table 2.2 Plasmids used in this study

Plasmid	Content	Source (ref.)
pUC19	Constuction vector	NEB
pJG4	9myc-MuB expressed from pET28a	(Ge et al., 2011)
pRHA-113	Construction vector, rhamnose-inducble promoter	(Giacalone et al., 2006)
pDB317	TetR-mCherry expression under control of salicylate-inducible promoter	(Joshi et al., 2013)
pGam-GFP	*gam-gfp* from SMR14334, inserted into Sal I/XbaI site of pRHA-113	This work
pGFP	*gfp* from SMR14334, inserted into Sal I/XbaI site of pRHA-113	This work
pKD3	Source of *cat*	(Datsenko and Wanner, 2000)

-Table 2.2 *continued*-

pKD4	Source of *kan*	(Datsenko and Wanner, 2000)
pKD13	Source of *kan*	(Datsenko and Wanner, 2000)
pRS306X112TetO	Source of $tetO_{112}$ array	(Michaelis et al., 1997)
pTetO(kan)	*kan* from pKD13 inserted into SalI /SpeI site of pRS306X112TetO	This work
pMuHF	~500 bp flanking the SE deletion inserted into EcoRI/SacI and SalI/HindIII sites of pUC19, respectively	This work
pTetO(kan)-MuHF	$tetO_{112}$*kan* cloned into SalI/SacI site of pMuHF	This work
pKD46	Plasmid for λ-Red recombination	(Datsenko and Wanner, 2000)
pCP20	Plasmid expressing Flp recombinase for antibiotics removal	(Datsenko and Wanner, 2000)

OLIGONUCLEOTIDE PRIMERS

Table 2.3 Oligonucleotide primers used in this study

Primer	Sequence (5’ -> 3’)
Insertion of *cat* into 35040 nt of Mu	
Owy185 (F)	gccagaagcctgatttaccgtttcctgtaaaccgaggttttggataatggggatccgtgtaggctg gagctgcttc
Owy186 (R)	ctggatctcctgtttaaaagcgccaatcatgccatgcgtgccaaaatcgaggatcccatatgaatat cctcctta
Substituting *SE* region with *cat*	
Owy061 (F)	atatttcaacgctgctgcgtaattaagaaggagaagaaattatgatggtgtaggctggagctgctt cg
Owy062 (R)	cgccgcgaaaaactgcaactgtcaaagatcatggaagacattatcacatatgaatatcctcctta
Monitoring integration of Mu DNA into host chromosome	
Mu_L1	gcttggttgttatcggttttgaacg
purH_R	cttcgacaaactgacgagaa
dnaC_F	ttgtgtcgcagtataccgc
dnaC_R	gaacttgacagccaaattcc
Sequencing Mu insertion site	
Mu_R1	gctacatcagattcctgaacaaacg
Mu_L1	gcttggttgttatcggttttgaacg
Detecting integrated Mu DNA and FD	
MuB (F)	gcaccacgccgtaaaggg
MuB (R)	ccttcttaattacgcagagc

-Table 2.3 *continued*-

MuR (F)	catttgaagcgcgaaagctaaag
lacZ (R)	caccgcgaggcggttttctccggc
IPE assay	
L1	gaacgttttttgaagctgttattgaa
L2	gcttggttgttatcggttttgaacg
R3	cgctacatcagattcctgaacaaacg
R4	ccaaactattcaaggttcagcca
Monitoring integration of Mu DNA into host chromosome	
L2	gcttggttgttatcggttttgaacg
purH (R)	cttcgacaaactgacgagaa
Quantifying integration frequency of Mu DNA	
Mu 1133 (F)	gtgcggatgtgattgcgggactt
Mu 1255 (R)	ctcacccttttggccagtgtcgtt
16S rRNA 101 (F)	gtggcggacgggtgagtaatgt
16S r RNA 202 (R)	cccctctttggtcttgcgacgtta
Gam-GFP construction	
P1	atcgtagtcgacatggctaaaccagcaaaacgtatcaag
P2	ccaatctagattatttgtatagttcatc
P3	gtagtcgacatggctagcaaag
Mu::TetO construction	
P4	attccggggatccgtcgacc

-Table 2.3 *continued*-

P5	cgactagtgtaggctggagctgcttcg
P6	gcgaattcttactgaactggcgtttg
P7	gcgagctcccatcataatttcttc
P8	gcgtcgactgataatgtcttccat
P9	cgaagcttgccatcggtaataaaacc
P10	aacctcaacgacgtttgcgccttctggattaattgctgaaacatccttttcaataattccggggatccgtcgacc
P11	tgatacacaacgcctgcgcgtccgcagcgttcatctgcaaaggactgaacccacttgtaggctggagctgcttcg

PHAGE

These were prepared by induction of the prophage strains by thermal inactivation of the temperature-sensitive (*ts*) phage repressor *c*, and concentrated by CsCl gradient centrifugation as described (Au et al., 2006). Prophages used in this study are listed in Table 2.1. For strains BU1717 (Mu*Bam*1066), SJ17 (Mu::Cm(*Bam*1066)) and SJ18 (Mu*Bam*1066ΔSE::Cm), the prophages were induced in the presence of pJG4 (c-myc MuB expressed from pET28(a) without IPTG induction) to supplement MuB protein. Phage from a 1 liter culture were precipitated with PEG 8000 and concentrated on a CsCl density gradient. Typical phage titers after concentration were ~10^{11} pfu (plaque forming units) / ml for wild type Mu, and ~10^{10} pfu for the *Bam* or *Bam*ΔSE phage, estimated by plating on BU1384 or BW25113. Phage titers for wild type Mu with and without the cat insertion were similar, showing that the insertion did not affect phage yields. Phage from the Mu::TetO lysogen were prepared in a smaller 50 ml culture and not concentrated on CsCl gradient. Instead, when cells reached an OD_{600} of 0.4, they were concentrated by pelleting and resuspensing in 10 ml of pre-warmed LB and induced as described above. Phage titers were ~10^{9} pfu / ml.

GROWTH CURVES

100 µl of saturated overnight cultures were transferred to 10 ml of fresh LB media and incubated at 37℃ until OD600 reached around 0.5 for all cultures. From then on, growth was monitored by measuring OD600 at various times for 2 hr. A similar procedure was followed for obtaining lytic growth curves, except that the LB media was

supplemented with 2.5 mM $CaCl_2$ and 5 mM $MgSO_4$. At OD600 of around 0.5, Mu phage was added to culture at MOI = 5, mixed briefly, and incubated at 37℃ for 3 hr until most cultures were completely lysed. In all cases where *priA2::kan, dnaT822* (without *dnaC* mutations) or *polA::kan* strains were used, these were grown overnight in minimal media, followed by dilution into fresh LB media, and then allowed to grow into log phase before infection with the different Mu phages.

PLAQUE MORPHOLOGY

10 μl of an appropriate dilution of phage suspension were mixed with 100 μl of host cells grown to 0.5–0.6 at OD600 in LB including 2.5 mM $CaCl_2$ and 5 mM $MgSO_4$. The mixture was added to 3 ml of 0.3% molten soft agar at 42℃, and poured on top of an LB agar plate containing 2.5 mM $CaCl_2$ and 5 mM $MgSO_4$. Plates were incubated overnight at 37℃.

HIGH-THROUGHPUT SCREENING OF THE KEIO LIBRARY

Cultures from the Keio collection stocked in 96-well plates were inoculated into new sterilized 96-well pates with 0.2 ml of Luria broth (LB) by using the 12-multichannel pipette (Biohit). They were incubated at 37°C overnight without shaking. 4 μl of saturated overnight cultures were transferred to 0.2 ml of fresh LB media supplemented with 2.5 mM $CaC1_2$ and 5 mM $MgSO_4$ in 96-well plates and incubated at 37°C until OD600 reached around 0.5, measured directly in the plates by DTX880 microplate reader (Beckman). Mu phage was added to the cultures at MOI = 5, mixed

briefly, and incubated at 30°C for 1 hr. 4 μl of infected cultures were spotted on slab agar plates having dimensions similar to the 96-well plates and containing 25 μg/ml chloramphenicol; plates were incubated overnight at 30°C.

LYSOGENIZATION/SURVIVAL FREQUENCY

Cultures were infected with Mu::Cm(*Bam*1066), Mu*Bam*1066ΔSE::Cm or Mu::Cm(*Aam*1093) phage as described under 'PCR-based assay for Mu integration'. Before and after infection, appropriate dilutions of cells in LB media were spread onto agar plates with or without 25 μg/ml chloramphenicol to obtain cell counts for input cells, survivors after infection, and lysogens. Plates were incubated at 30°C overnight, and colonies were counted the next day. Lysogenization efficiency was calculated as Cm^R cells/input cells x 100, and survival efficiency was calculated as survivors (on non-antibiotic plate) / input cells x 100.

BLOCKING REPLICATION IN DNA*TS* MUTANTS

For all Dna*ts* mutants, cells grown at 30°C were shifted to 42°C for variable times, depending on the mutation. DnaA*ts* and DnaC*ts* mutants were incubated for 90 min at 42°C to allow ongoing replication to terminate (Wechsler and Gross, 1971), while all other *ts* mutants were incubated for 30 min at 42°C, because these mutants stop replication immediately at the non-permissive temperature (Saluja and Godson, 1995).

VISUALIZING MU AND DSBS *IN VIVO*

Mu::TetO location was visualized by expressing TetR-mCherry from pDB317 using 100 μM sodium salicylate. Double strand breaks (DSBs) were assayed by appearance of either Gam-GFP foci (Shee et al., 2013) or RecA-GFP foci (Renzette et al., 2005). Gam-GFP expression protocol was modified slightly from that of Shee et al. who expressed it for 4 hr to visualization (Shee et al., 2013). Prolonged expression of Gam-GFP affects cell viability (Fig. 4.10). Therefore, expression was controlled by the tightly regulated rhamnose-inducible promoter in pGam-GFP by repressing it with 0.2% glucose until needed. 500 μM rhamnose was added only 30 min before Mu infection. RecA-GFP is expressed constitutively from its normal chromosomal location in the *recA-gfp* strain (SJ006).

Mu::TetO infection was carried out at an MOI of 1 in host strains already expressing TetR-mCherry. 15 min after infection, cells were placed onto 1% agarose pads at room temperature as described (Skinner et al., 2013), and visualized within 5-10 min with an Olympus BX53 fluorescence microscope. Images were captured using cellSens standard software (version 1.6) from Olympus.

MEASURING THE SOS RESPONSE USING GFP

This assay measures GFP expression from the promoter of the SOS-induced gene *sulA* (*PsulA-gfp*; (McCool et al., 2004b). Mu infections were carried out either in WT (SS996), or its *lexA3* (SS4294) and *lexA71* (SS4610) derivatives. *lexA3* is defective for SOS induction, while *lexA71* is constitutively induced. 30 min after Mu infection, cells

were harvested and diluted 1:100 into TE buffer. Cells were sorted in the BD Accuri™ flow cytometer and analyzed as described above.

PCR-BASED ASSAY FOR MU DNA INTEGRATION

50 µl overnight cultures were transferred to 5 ml of fresh LB media supplemented with 2.5 mM $CaC1_2$ and 5 mM $MgSO_4$ and grown to 0.5 at an OD600. Phage were added to the cultures at MOI = 5 and incubated at 30℃ for 30 min. Infected cells were harvested and the total DNA were isolated by Wizard Genomic DNA purification kit (Promega). PCR was conducted with 50 ng DNA as a template, 10 pmol primers, 1X *Go Taq*™ master mix (Promega), and distilled water up to 50 µl. Primers were designed to anneal to the left end of Mu DNA and the *purH* gene of *E. coli*. PCR conditions were: 94℃ for 2 min, 30 cycles of - 94℃ for 30 sec, 50℃ for 30 sec, 72℃ for 2 min 30 sec - and a final extension at 72℃ for 2 min. PCR amplification primers used in this study are listed in Table 2.3. The reaction products were electrophoresed on 1% agarose gels and visualized by staining with ethidium bromide.

QUANTITATIVE REAL-TIME PCR ANALYSIS

This method measures DNA amounts based on the fluorescence signal from SYBR-bound DNA. PCR reactions were conducted with the same templates and primers as used for normal PCR, with the additional inclusion of 1X Power SYBR Green PCR Master Mix (Applied Biosystems), and distilled water up to 25 µl. The PCR program in the 7900HT or ViiA7 sequence detector (Applied Biosystems) was as follows: 95℃ for

10 min, followed by several cycles of - 95℃ for 30 sec, 50℃ for 30 sec, 72℃ for 2 min 30 sec. Cumulative fluorescence was measured at the beginning of the exponential phase of the PCR reaction to determine the fractional cycle number (C_T). The level of integrated Mu DNA was normalized to a chromosomal locus *dnaC* or 16 rRNA amplified with appropriate primers listed in Table 2.3.

ITERATIVE PRIMER EXTENSION (IPE) ASSAY

This assay has been described (Pato, 2004). One-directional PCR was performed with 1μg of total genomic DNA as a template, 5 pmol of a single primer which was 5'end-labeled with γ-^{32}P, Taq 2X master mix solution (Promega). Primer was extended as follows: 94°C for 2 min; 60 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 20 s; and a final extension at 72 for 7 min. PCR products were separated on a 10% polyacrylamide sequencing gel and quantified with a phosphorimager, Typhoon 9500 (GE Life Science).

DETECTING INTEGRATED MU AND ITS FLANKING DNA IN THE *E. COLI* GENOME BY PCR

200 μl of overnight cultures were transferred into 20 ml LB media (5 mM $CaCl_2$ and 5 mM $MgSO_4$) and grown till an OD_{600} of ~ 0.3-0.4. All infections with wild type Mu phage were at MOI = 5 at either 30°C, 37°C or 42°C. Mu infections in the Dna*ts* strains were carried out at 42°C. At various times after infection, total DNA was extracted, subjected to pulse field gel electrophoresis, and the gDNA band isolated as described (Au et al., 2006). Mu integration and flanking DNA (FD) sequences were amplified by standard PCR as described, and the products visualized on 1% agarose gels

after staining with ethidium bromide (Choi and Harshey, 2010). Mu integration was detected by PCR using primers within the MuB gene, and FD DNA by primers that amplified the junction between right end of Mu and *lacZ* (Table 2.3). PCR was performed with 50–100 ng of template DNA, 10 pmol of primers, 10 µmol of deoxynucleoside triphosphates, 2.5 units of Taq polymerase (Qiagen), 1x PCR buffer, and 1x Q solution in 25 µl. The PCR conditions were 94°C for 2 min; 30–40 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 30 s; and a final extension at 72°C for 7 min as described (Choi and Harshey, 2010). SS996 derivative strains include a partial *lacZ* gene, but this did not interfere with the assay because amplification was based on a second primer annealing inside Mu.

SEQUENCING MU INSERTION SITES IN THE *PRIA* MUTANT

priA lysogens were selected as Cm^R colonies after infection with Mu::Cm(*Bam*1066) phage. After overnight culture into LB media, chromosomal DNA was isolated by Wizard Genomic DNA purification kit and digested by restriction enzyme BamH I and Pst I. Digested DNA fragments were purified and ligated with similarly digested pUC19 plasmid. Cm^R transformants were isolated and digested by BamH I and Pst I to ascertain that the insert size was larger than 4 kb, so that it included DNA flanking the insertion. R1 primer (Table 2.3) was annealed to Mu DNA right end to obtain sequence of the flanking DNA. Based on this sequence, appropriate primers were used to PCR-amplify DNA flanking the left end of the insertion using the L1 primer. DNA sequencing was performed at our core sequencing facility.

FLOW CYTOMETRY

Flow cytometry was used to determine the number of chromosomal origins per cell. 1 ml culture samples were fixed by adding 9 ml of 95% ethanol and maintained at 4°C. These cultures were centrifuged at 6000g for 10 min in 4°C and the cell pellet was washed with and resuspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Cells were diluted to OD_{600} of 0.1 and stained by addition of 5 μM SYTOX Green (Invitrogen) by incubating in the dark for 15 min. A total of ~20,000 viable cells were sorted in a BD Accuri™ flow cytometer, and the data were analyzed by FlowJo software.

WESTERN BLOT

After Mu infection for 30 min, 5 x 10^8 cells were harvested and lysed in lysis buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA and 0.02 % bromophenol blue). After boiling at 95°C for 5 min, samples were applied to 10% SDS-PAGE gel, transferred to a PVDF membrane (Bio-Rad) after electrophoresis, and the blot was probed with a polyclonal anti-MuB antibody (Parsons and Harshey, 1988), followed by HRP-conjugated goat anti-rabbit IgG (Bio-Rad), and detected using ECL western blotting analysis reagents (GE Healthcare) (Ausubel and *al*, 2003).

Chapter 3. The double-strand break repair pathway of *Escherichia coli* is required to recover Mu insertions

Portions of this chapter have been published in PLoS Genetics (2012), 8(4): e1002642, Jang S, Sandler SJ, Harshey RM. "Mu insertions are repaired by the double-strand break repair pathway of Escherichia coli."

ABSTRACT

Mu is both a transposable element and a temperate bacteriophage. During lytic growth, it amplifies its genome by replicative transposition. During infection, it integrates into the *Escherichia coli* chromosome through a mechanism not requiring extensive DNA replication. In the latter pathway, the transposition intermediate is repaired by transposase mediated resecting of the 5' flaps attached to the ends of the incoming Mu genome, followed by filling the remaining 5 bp gaps at each end of the Mu insertion. It is widely assumed that the gaps are repaired by a gap-filling host polymerase. Using the *E. coli* Keio Collection to screen for mutants defective in recovery of stable Mu insertions, we show in this study that the gaps are repaired by the machinery responsible for the repair of double-strand breaks in *E. coli*—the replication restart proteins PriA-DnaT and homologous recombination proteins RecA, RecB, and RecC. We discuss alternate models for recombinational repair of the Mu gaps.

INTRODUCTION

Transposable elements drive genome evolution in many ways – increasing DNA content, rearranging and mutating genes, as well as altering gene regulation (Craig, 2002). Temperate phage Mu has played a pivotal role in our current understanding of how movable elements move (Symonds et al., 1987). A unique aspect of Mu is that, depending on the phase of its life cycle, it moves using either replicative or non-replicative modes of DNA transposition (Chaconas and Harshey, 2002). Most of our knowledge of Mu transposition is derived for the replicative pathway, where during lytic growth, Mu amplifies its genome by repeated transposition replication events which exploit the host replication apparatus (Mizuuchi, 1992; Nakai et al., 2001). *In vitro* experiments have established that in this pathway, the Mu transposase (MuA protein) mediates single-strand cleavages at Mu ends followed by strand transfer of the cleaved ends into target DNA; the latter reaction is greatly assisted by MuB protein (Fig. 1.6). The resulting branched strand transfer joint is resolved by target-primed replication, which is initiated by the PriA primosome and completed by the Pol III holoenzyme, and results in duplication of the Mu genome after every round of integration. At the end of the lytic cycle, Mu genomes are packaged into phage heads such that they include host sequences (flaps) from both sides of a Mu insertion.

The non-replicative pathway of Mu transposition is only used when progeny phage infect new hosts (Liebart et al., 1982; Akroyd and Symonds, 1983; Harshey, 1984). Along with Mu DNA, the phage also inject into the host the phage N protein, which binds at the termini and converts the linear Mu genome into a non-covalently closed supercoiled circle (Harshey and Bukhari, 1983; Puspurs et al., 1983; Gloor and Chaconas,

1986). Integration of the infecting Mu into the host genome follows the same initial nick-join steps of transposition established for the replicative mechanism *in vitro*; however, instead of target-primed Mu replication, the host flaps are resected and the gaps are repaired (Fig. 1.6). Flap resection has been demonstrated both *in vivo* and *vitro*. This reaction is dependent *in vivo* on the cryptic endonuclease activity harbored within the C-terminal domain of the transposase MuA (designated MuA_{Nuc} in this study), as well as on the chaperone protein ClpX (Wu and Chaconas, 1995; Choi and Harshey, 2010). ClpX is known to play an essential role during Mu replication, remodeling the Mu transpososome and enabling its transition to a replisome (Nakai et al., 2001; Abdelhakim et al., 2008) (Fig. 1.6). The alternative choices for resolving the transposition intermediate, i.e. repair versus replication, must involve additional phage and host factors whose identity is not yet established.

The current study was undertaken to identify host factors involved in the repair of Mu insertions during the non-replicative infection pathway. To do so we used the Keio Collection, which is a set of 3,985 precisely defined, single-gene deletions of all nonessential genes in *Escherichia coli* K-12 (Baba et al., 2006), and screened for mutants defective in recovery of Mu::Cm insertions. Among the several mutants that gave a poor yield of CmR integrants, a majority of those that allowed Mu entry showed normal integration and replication of wild type Mu. By using two additional phage variants to re-screen/re-test in order to eliminate those defective in maintenance of a stable prophage state, we narrowed the search to a small subset of the mutants. Included among these were mutants in the homologous recombination pathway - *recA, recB, recC*. Two mutants - *priA* and *dnaT* – were defective in Mu replication as expected, but were unexpectedly defective in the recovery of insertions despite being proficient in Mu

integration. The data show that Mu insertions are repaired by the replication restart machinery and homologous recombination proteins.

RESULTS

A functional map of the Mu genome is shown in Figure 3.1A. A ~1 kb *cat* cassette encoding chloramphenicol (Cm) resistance was inserted into a non-essential region of the prophage genome (see Materials and Methods). Phage derived from this strain were used to infect the Keio mutant collection (see Table 2.1 for strain information), which occupies forty-eight 96-well plates, and spotted on agar slabs containing chloramphenicol to select for Mu lysogens as described in Materials and Methods. The control panel in Figure 3.1B shows results expected for known hosts where Mu integrates, but either does or does not replicate. In our standard wild type host BU1384 where Mu replicates, ~90% of the infected cells undergo lytic growth and lysis, and ~10% of the survivors (i.e. ~1% of input cells) are lysogens. Mu fails to replicate in isogenic strains carrying either a *himA* or a *clpX* null mutant allele. *himA (ihfA)* codes for one of the two subunits of the regulatory protein IHF, which is required for early Mu gene transcription (Symonds et al., 1987; Higgins et al., 1989), and ClpX is essential for Mu replication (Mhammedi-Alaoui et al., 1994; Nakai et al., 2001). Both of these mutant strains support Mu integration (Chaconas et al., 1984; Au et al., 2006; Choi and Harshey, 2010). A larger number of Cm^R colonies are recovered in these strains compared to wild type because Mu does not undergo lytic development. Similar differences in the recovery of Cm^R colonies were seen in the wild type Keio strain BW25113 and its isogenic *himA* and *clpX* derivatives. In our screen for repair-defective mutants, we expected to identify mutant spots with either no Cm^R colonies or with fewer colonies than wild-type.

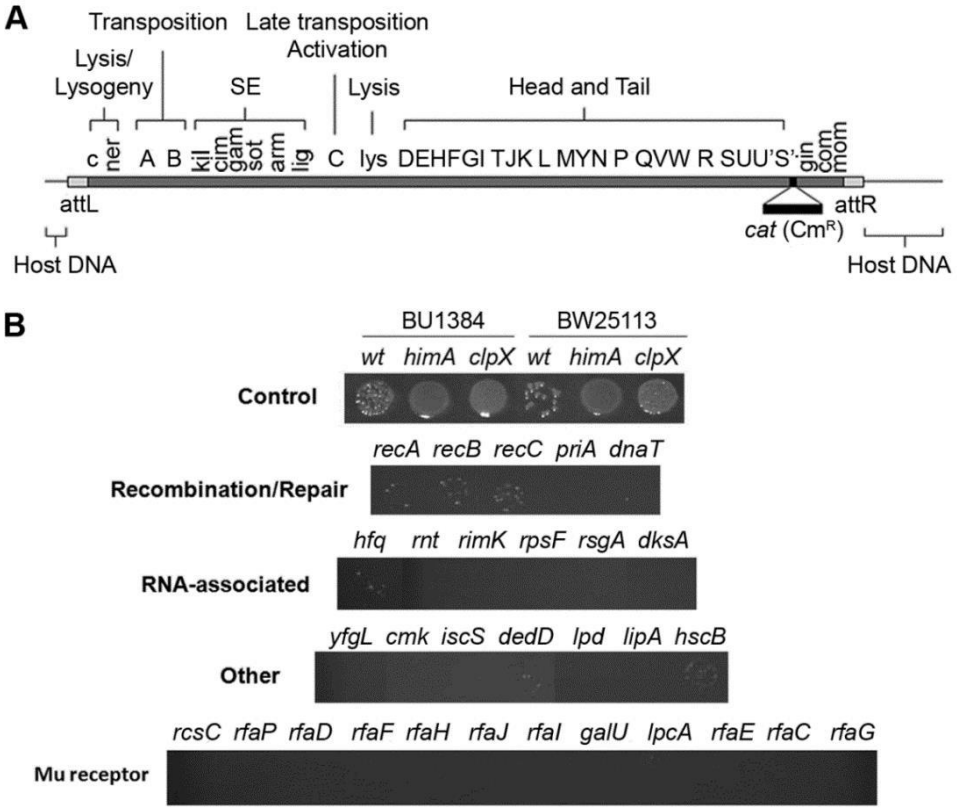


Figure 3.1 Identification of *E. coli* mutants in the Keio library defective in recovery of Mu::Cm insertions.

(**A**) Functional map of the Mu genome packaged within a phage particle, showing position of inserted CmR cassette, and host or flap DNA attached to both ends. The SE (semi-essential) region contains 14 orfs (Morgan et al., 2002); only those assigned a phenotype/function are indicated (Symonds et al., 1987). (**B**) Cultures infected with Mu::Cm were spotted on Cm plates as described in Materials and Methods. Control panel: Expected results from infection of two different wild type and their derivative mutant strains - *clpX* and *himA* - that do not support Mu replication. Bottom four panels: Final set of mutants from the Keio screen showing lower Mu::Cm lysogen recovery compared to the wild type strain, grouped into indicated categories.

The majority of mutant strains behaved like wild type in this screen. Known host mutants that do not support replication were easily identified (Fig. 3.2, see plate #1), but no new candidates with this phenotype were observed. Several mutants displayed the phenotype of interest i.e. showed fewer or no colonies in the spots compared to wild type (Fig. 3.2, see plate #1 and #9). The phenotype of these latter mutants was re-confirmed by infecting with Mu phage carrying a different antibiotic resistance marker (Mu::Amp) to ensure that the phenotype was independent of the antibiotic used for selection. The final set of 30 mutants displaying this phenotype is arranged in four panels below the control panel in Figure 3.1B. The mutants are classified broadly into genes known to affect DNA recombination/Repair, RNA-associated functions, 'Other' functions, and Mu receptor function. A more detailed description of gene function is listed in Table 3.1.

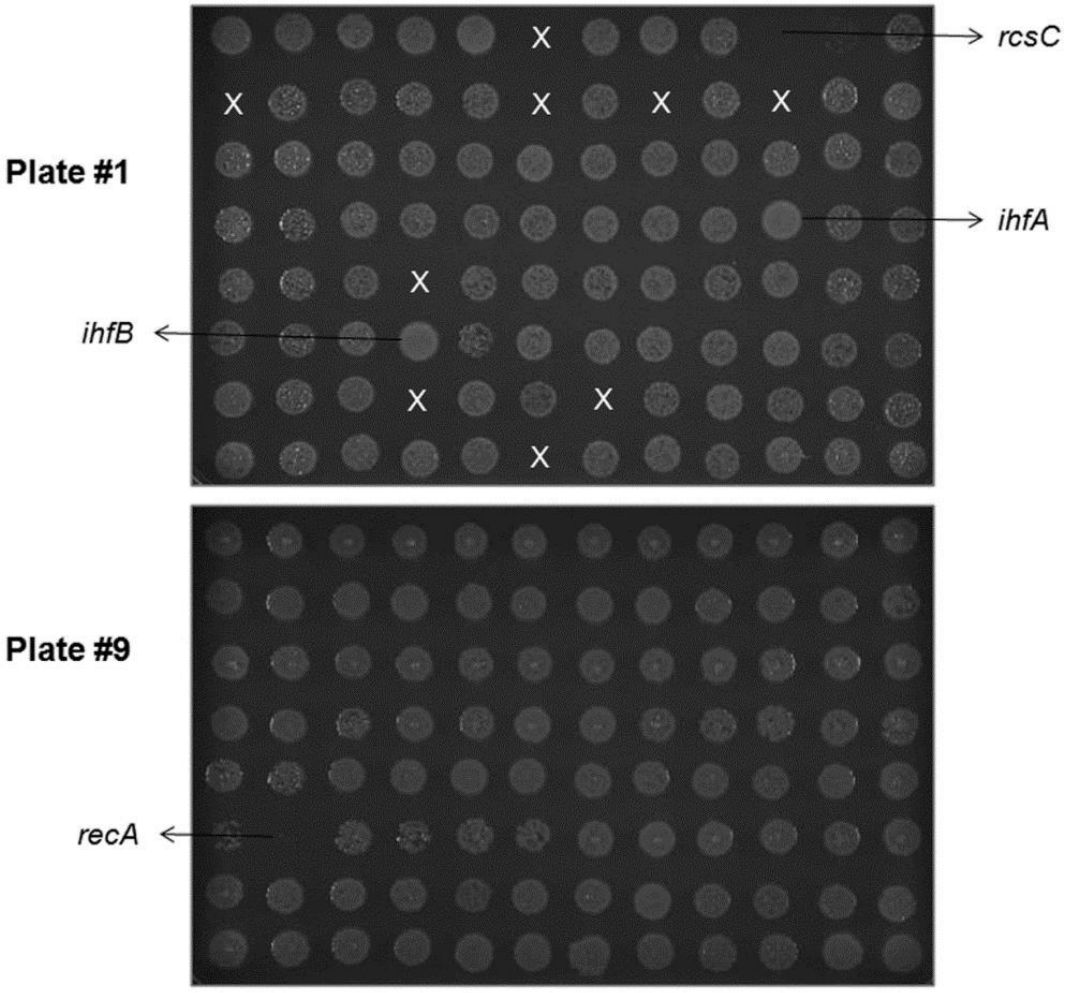


Figure 3.2 Initial results of spotting Mu-infected cultures derived from Keio plates. Samples of #1 and #9 on LB Cm plates. X marks spots with no bacteria.

Table 3.1 Description of mutants defective in Mu lysogen recovery

(**A**) Defect in lysogen recovery

No.	ID (Keio)	Gene	Function
GROUP 1. Repair/Recombination associated proteins			
1	2669	*recA*	DNA strand exchange and recombination protein with protease and nuclease activity
2	2788	*recB*	Exonuclease V (RecBCD complex), beta subunit
3	2790	*recC*	Exonuclease V (RecBCD complex), gamma chain
4	3906	*priA*	Primosome factor n' (replication factor Y)
5	4326	*dnaT*	DNA biosynthesis protein (primosomal protein I)
GROUP 2. Ribosomal RNA associated proteins			
6	141	*dksA*	Transcriptional regulator of rRNA transcription, DnaK suppressor protein
7	4130	*hfq*	RNA-binding protein that affects many cellular processes
8	1644	*rnt*	Ribonuclease T (RNase T)
9	836	*rimK*	Ribosomal protein L6 modification protein
10	4158	*rpsF*	30S ribosomal subunit protein S6
11	4122	*rsgA*	Ribosome small subunit-dependent GTPase A
GROUP 3. No category (Other)			
12	2496	*yfgL*	protein assembly complex, lipoprotein component
13	2511	*hscB*	Hsc20 co-chaperone that acts with Hsc66 in IscU iron-sulfur cluster assembly
14	893	*cmk*	Cytidylate kinase
15	2514	*iscS*	Cysteine desulfurase (tRNA sulfurtransferase), PLP-dependent
16	112	*lpd*	Lipoamide dehydrogenase, NADH-dependent
17	623	*lipA*	Lipid synthesis, iron-sulfur protein
18	5378	*dedD*	Membrane-anchored periplasmic protein involved in septation

-Table 3.1 *continued*-

(**B**) Defect in Mu DNA entry

No.	ID (Keio)	Gene	Function
1	1224	*galU*	glucose-1-phosphate uridylyltransferase
2	212	*lpcA*	D-sedoheptulose 7-phosphate isomerase
3	5917	*rcsC*	hybrid sensory kinase in two-component regulatory system with RcsB and YojN
4	3596	*rfaC*	ADP-heptose:LPS heptosyl transferase I
5	3594	*rfaD*	ADP-L-glycero-D-mannoheptose-6-epimerase, NAD(P)-binding
6	3024	*rfaE*	fused heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase
7	3595	*rfaF*	ADP-heptose:LPS heptosyltransferase II
8	3606	*rfaG*	glucosyltransferase I
9	3818	*rfaH*	DNA-binding transcriptional antiterminator
10	3602	*rfaI*	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase
11	3601	*rfaJ*	UDP-D-glucose:(galactosyl)lipopolysaccharide glucosyltransferase
12	3605	*rfaP*	kinase that phosphorylates core heptose of lipopolysaccharide

ID numbers and associated gene descriptions are from the Keio web site, www.ecolicommunity.org/genobase.

Mu integration and replication in *E. coli* mutants defective in lysogen recovery

The poor yield of Cm^R colonies in the mutants shown in Figure 3.1B could be due to defects in Mu entry, integration, stable maintenance of lysogeny, or repair. To distinguish between some of these possibilities a PCR assay was first employed to test for Mu integration (Fig. 3.3A). Two primers were chosen to amplify covalent junctions between the left end of Mu DNA and an arbitrarily chosen target gene *purH*. A PCR product is expected once the 3' ends of Mu are joined to the target regardless of the fate of 5' ends (see Fig. 1.6). PCR products of different lengths are expected since Mu integration is essentially random (Manna et al., 2004; Ge et al., 2011). Using this method, a control experiment first followed the time course of wild type as well as mutant *Bam* and *Aam* Mu phage infections in the wild type strain. The particular *Bam* mutation used here (*Bam*1066) is reported to be fairly proficient in integration but defective in replicative transposition of Mu (Chaconas et al., 1985). The *Aam* mutant (*Aam*1093) is defective in integration (O'Day et al., 1978). The integration patterns obtained during these infection experiments were consistent with the known transposition properties of these phages (Fig. 3.3A).

Wild type Mu was used to infect the 30 mutants obtained in the initial screen for repair-defective mutants (Fig. 3.1B). Mutants grouped under Recombination-Repair, RNA and Other categories all showed similar levels as well as patterns of integration compared to the wild type strain (Fig. 3.3B). Quantitative PCR with a subset of these mutants (*priA, recA*) validated the results with normal PCR (Fig. 3.4; we note that Southern blots used in earlier studies also showed similar levels of Mu integration in wild type and *priA* mutants (Jones and Nakai, 1997)). Thus, these mutants were not defective in either Mu entry or integration. A majority of the mutants with defects in the LPS

biosynthesis pathway, however, showed little or no integration (Fig. 3.3C). This is likely due to a block in Mu entry, since the receptor for Mu is located within the LPS (Sandulache et al., 1984; Muller et al., 1988).

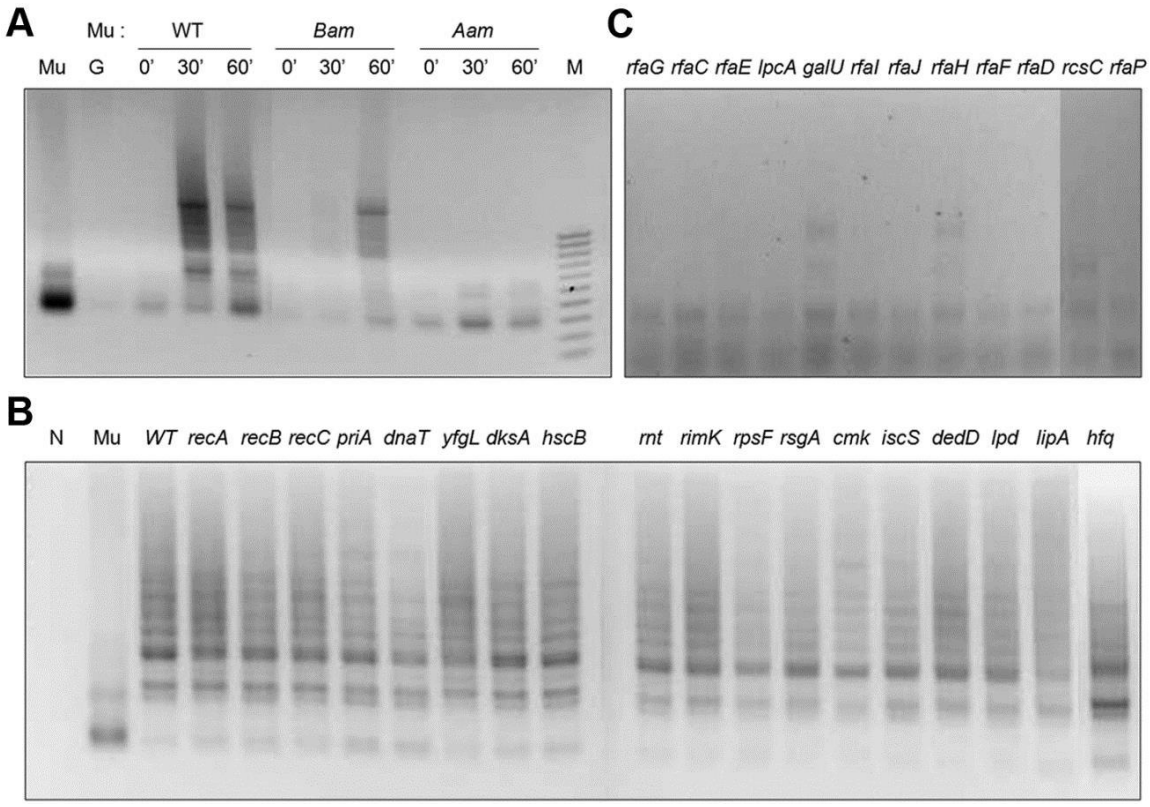


Figure 3.3 PCR assay for Mu integration in mutants defective in lysogen recovery.

(**A**) Control PCR reactions monitoring integration at different time points after infection of wild type BW25113 with Mu::Cm, Mu::Cm(*Bam*1066) and Mu::Cm(*Aam*1093) phage. These phages can integrate-replicate, integrate but not replicate, or not integrate, respectively. (**B**) PCR results for wild type Mu::Cm integration 30 min after infection of mutants in the first three categories shown in Figure 3.1B. **(C)** As in (B) but with mutants in the Mu Receptor category. Control reactions with either no template (N), Mu, or genomic DNA templates from uninfected BW25113 host (G) are indicated, along with size markers (M). Reaction products were run on agarose gels and stained with ethidium bromide as described under Materials and Methods.

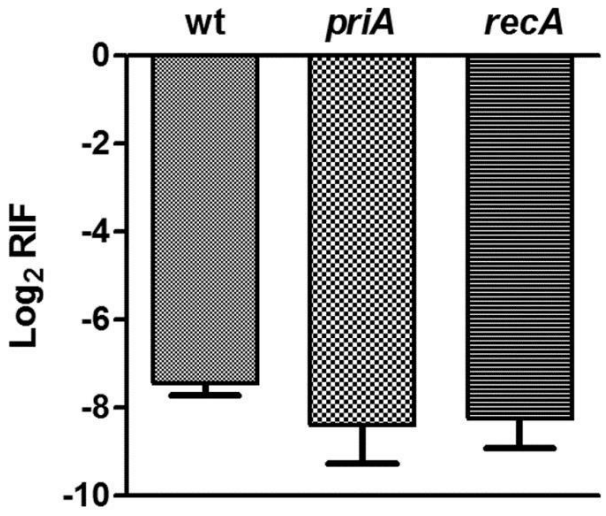


Figure 3.4 Quantitation of Mu DNA integration in wild type, *priA* and *recA* mutant strains by real-time PCR analysis.

Genomic DNA isolated from the indicated Mu-infected strains was used in real-time quantitative PCR reactions to quantify Mu integration as described in Materials and Methods. C_T is the fractional cycle number at the beginning of the exponential reaction phase where the fluorescence passes a threshold (T) at which the fluorescence signal is first detected. C_T values are inversely proportional to the amount of amplified DNA. ΔC_T = Mu C_T – dnaC C_T. dnaC is used as a control for as a single-copy chromosomal gene. The data are an average of three technical repeats.

To test if mutants that supported integration also supported Mu replication, cell lysis and phage production were monitored. Growth of the strains with and without Mu infection is shown in Figure 3.5. The LPS mutants in Figure 3.1B all grew as well as wild type; only a representative mutant *rfaF* is shown in Figure 3.5A. Neither this mutant, nor others in this category were susceptible to lysis by Mu infection (Fig. 3.5B), supporting the conclusion that this group of mutants is defective in Mu entry. They were therefore not studied further. The remaining mutants showed varying degrees of growth impairment compared to wild type (Fig. 3.5A). With the exception of *priA* and *dnaT*, which are essential for Mu replication (Nakai et al., 2001), cell lysis and phage production were observed in all of the infected strains (Fig. 3.5B). Thus, the majority of these mutants supported both Mu integration and replication. Their defect in yielding stable lysogens could therefore be due to an inability to maintain lysogeny or defects in repair of the insertions.

Defects in maintenance of the prophage state or lysogeny might be discerned by examining Mu plaque morphologies on these mutants. These would be expected to have a 'clear' rather than the 'turbid' phenotype observed for wild type Mu, which can be maintained in a lysogenic state. *dksA, hfq, rnt and rpsF* gave turbid plaque morphologies somewhat similar to the wild type strain, *dedD* was apparently clear, while the remaining mutants had clear centers and clear edges with turbid rings in-between (Fig. 3.6). In the latter set of mutants with the mixed clear-turbid phenotype, it was difficult to ascertain whether the lysogeny-maintenance function might be affected.

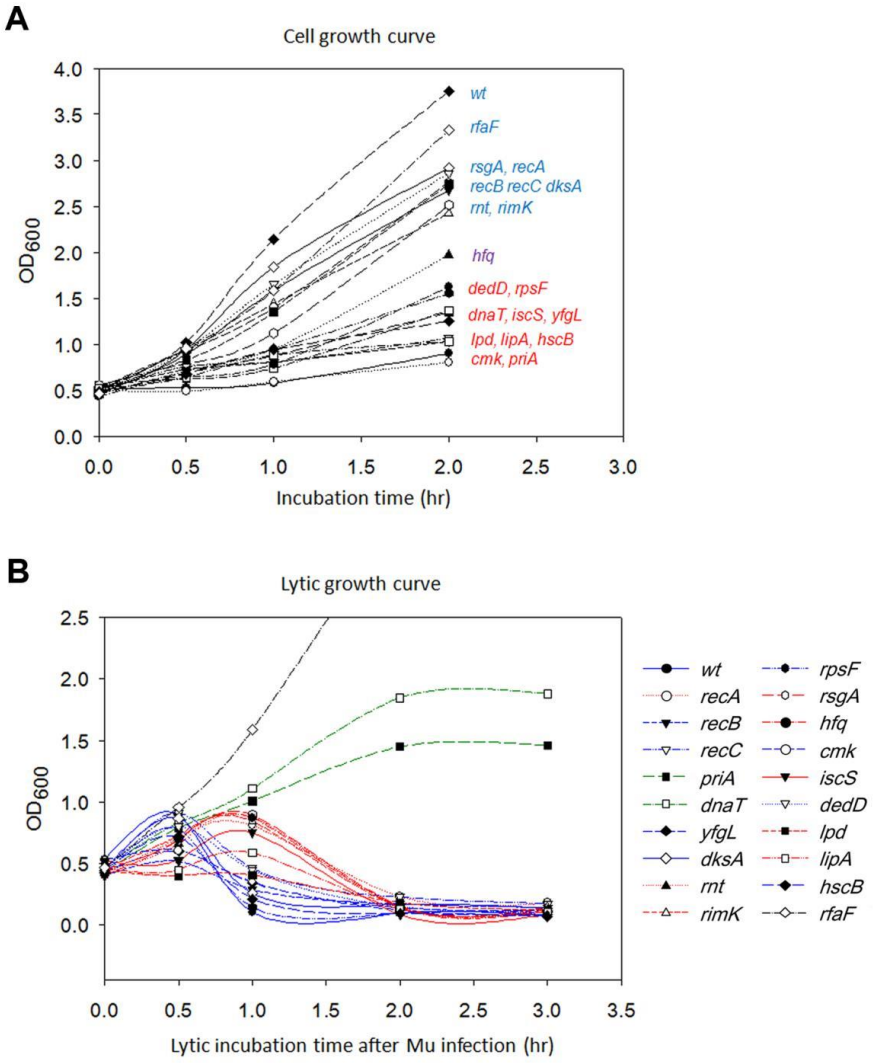


Figure 3.5 Mu replication in mutants defective in lysogen recovery.

(**A**) Growth curves of mutants, color coded to indicate slow (red), medium (purple) or near-wild type (blue) patterns. (**B**) Lysis profile of mutants after infection with wild type Mu, color coded to indicate similarity to wild type (blue), slightly delayed from wild type (red), growth delay but no lysis (green), and no lysis (black). All strains were grown to OD_{600} of ~0.5 prior before infection with Mu::Cm. Phage production in the lysed cultures was monitored by titration (not shown).

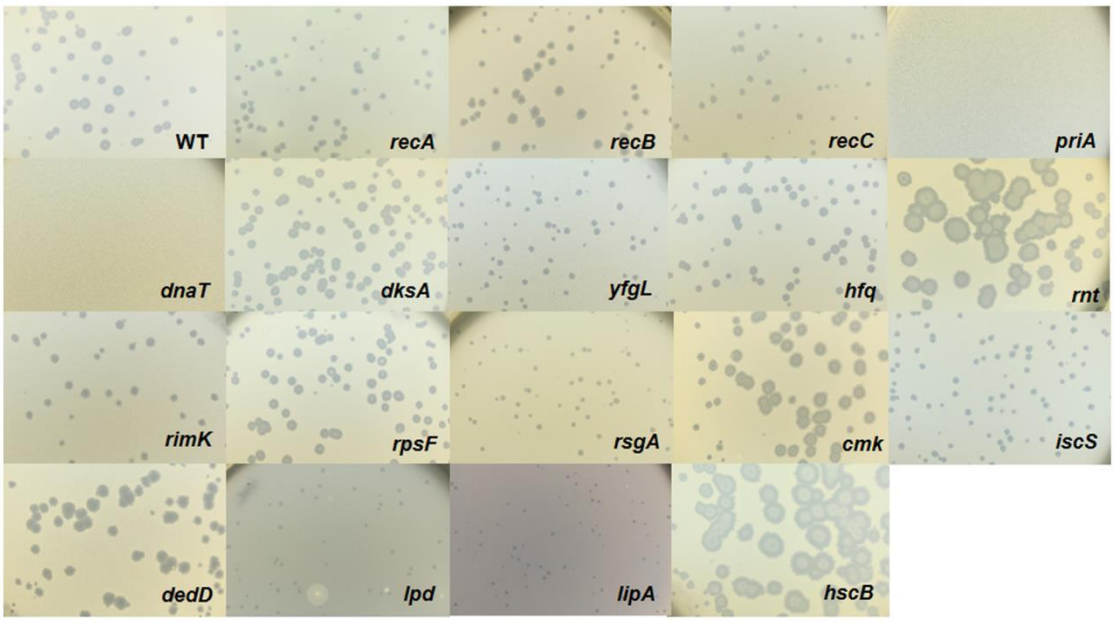


Figure 3.6 Plaque morphologies of wild type Mu::Cm on Keio mutant strains defective in lysogen recovery.

Keio mutant screen with replication-defective Mu

To eliminate scoring mutants as repair-defective because they were unable to maintain the lysogenic state and were therefore going lytic, we re-screened the Keio library with a Mu::Cm variant defective in replication. This phage carries the *Bam*1066 mutation, which allows integration but does not support replicative transposition (see Fig. 3.3A; (Chaconas et al., 1985)). The same set of mutants was isolated in this screen as well. In the spot test results shown in Figure 3.7, it appears that some of the mutants have more Cm^R colonies than obtained with wild type phage (see Fig. 3.1B). This is because a higher proportion of cells survive during infection with this phage due to absence of lytic growth. Lysogen recovery was therefore quantified as described under Materials and Methods (Fig. 3.8A). Among mutants in the Recombination-Repair category, *priA* and *dnaT* mutants were the most severely affected in lysogen recovery (0.04%), followed by *recA* (0.2%), *recB* (0.7%) and *recC* (0.9%). Among mutants in the RNA and Other category, with the exception of *yfgL*, *dksA*, *hfq*, *rimK* and *lpd*, the remainder had lysogen frequencies similar to or even better than wild type.

A surprising aspect of the data shown in Figure 3.8A is that lysogen recovery in the wild type was only ~5% with Mu*Bam* phage, and that cell viability after infection was only ~20% (Fig. 3.9A). Similar low cell viability was observed even after infection with integration-defective Mu*Aam* phage (Fig. 3.3A and Fig. 3.9A). To test if this was due to expression of the cell killing function *kil* or to other function(s) specified by the unknown orfs in the SE (semi-essential) region (Morgan et al., 2002), which is transcribed as part of a long early transcript that includes the *A* and *B* genes (Symonds et al., 1987) (see Fig. 3.1A), we deleted the SE region in the Mu*Bam* phage (see Materials and Methods). Indeed, infection with Mu*Bam*1066ΔSE::Cm phage improved both lysogen recovery and

cell viability in the wild type to 100% (Fig. 3.8B and Fig. 3.9B, respectively). Under these conditions, all the mutants in the Recombination-Repair category still remained impaired (<15% of wild type) for lysogen recovery. In the RNA/Other category, *hfq, lpd* and *lipA* were also still substantially impaired (18-25% of wild type). Since *hfq* shows wild type plaque morphology (Fig. 3.6) and since there is no obvious relationship of the known functions of these three genes to DNA repair, we will not consider them further here.

We conclude that a majority of the *E. coli* genes required for recovery of stable Mu insertions provide functions that apparently allow host survival in the presence of lethal phage functions specified by the SE region of Mu. The group of five genes that remain defective - *priA, dnaT, recA, recB* and *recC* – is significant in that this group is known to participate in recombinational repair. The isolation of this group of genes must be related to the repair of Mu insertions and not to repair of random double strand breaks generated upon Mu infection, because (1) they are dependent on Mu integration (i.e. infection with Mu*Aam*1093 phage does not significantly affect the viability of the *priA* and *recA* hosts as compared to wild type; Figure 3.9A, and (2) Mu-induced mutations are known to be tightly linked to Mu i.e. they are not random (Taylor, 1963).

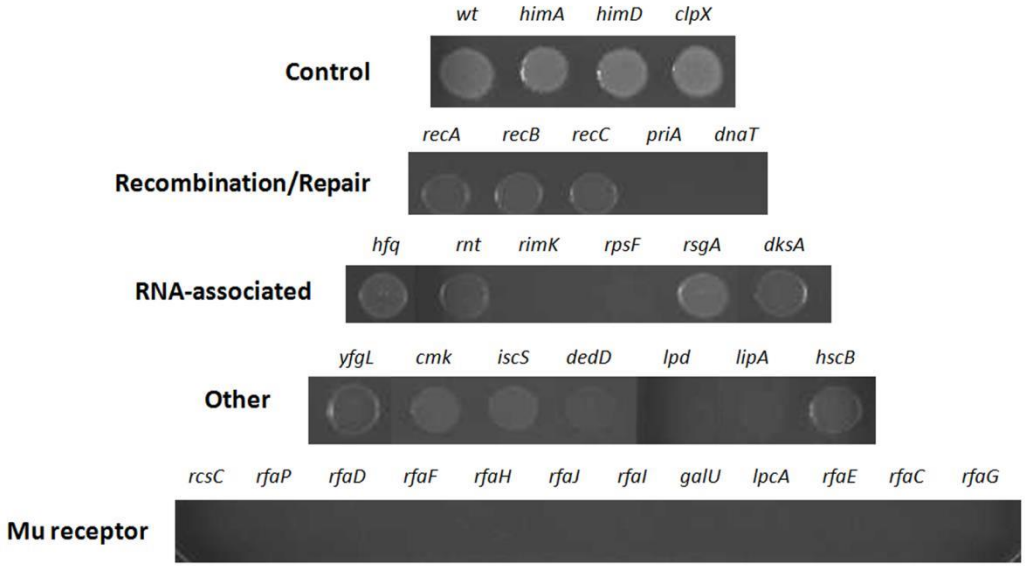


Figure 3.7 Keio mutant screen using Mu::Cm(*Bam*1066).

Final set of mutants obtained are shown. Spot tests and mutant categories are as in Figure 3.1B, except that strains in the control panel are all derived from BW25113. *himA* (*ihfA*) and *himD* (*ihfB*) code for the two subunits of IHF, which is essential for the Mu replicative pathway.

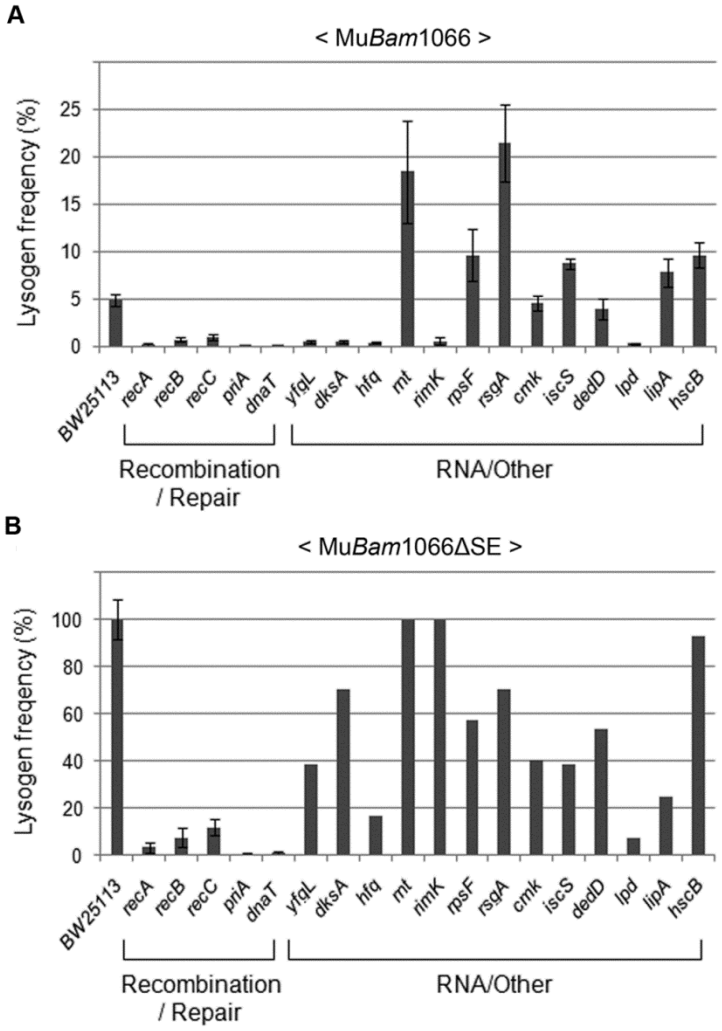


Figure 3.8 Mutant screen using replication-defective Mu.

Lysogenization efficiencies (calculated as Cm^R cells/infected cells x 100), of the mutant strains infected with either (**A**) Mu::Cm(*Bam*1066) or (**B**) Mu*Bam*1066ΔSE::Cm. Mutant categories as in Figure 3.1B. Error bars indicate standard deviation from the mean of triplicate data sets obtained from three independent colonies of the same strain. In (B), data for RNA/Other mutants are from a single colony/experiment. See Materials and Methods for details.

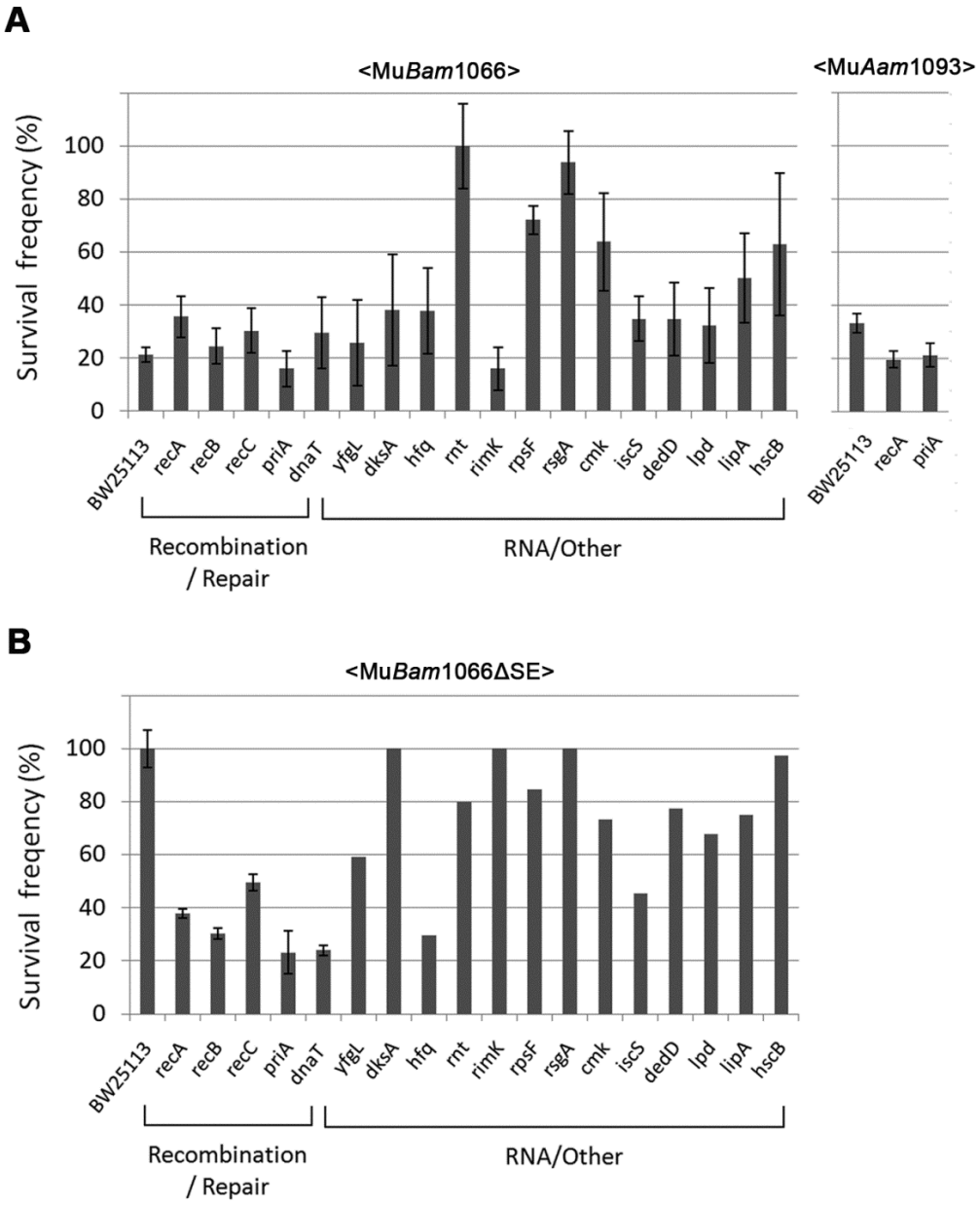


Figure 3.9 Survival efficiency of mutant strains infected with (**A**) Mu::Cm(*Bam*1066) and Mu::Cm(*Aam*1093) or (**B**) Mu*Bam*1066ΔSE::Cm phage.

Survival efficiency is calculated as cells recovered after infection on no-antibiotic plates/infected cells×100. See Materials and Methods and Figure 3.8 legend for other details.

Role of replication restart in the non-replicative pathway of Mu transposition

PriA and DnaT play a central role in the repair of nicks and gaps created by DNA damaging agents in *E. coli* by promoting replication restart after fork collapse, either with or without the involvement of recombination (Gabbai and Marians, 2010). There are multiple pathways for replication restart that require PriA, PriB, PriC, DnaT and Rep (Gabbai and Marians, 2010). These proteins identify the correct substrate, process it if necessary, and then aid DnaC in loading the replicative helicase DnaB during pre-primosome formation. PriA and DnaT are required for the two main pathways of 'Restart' where PriB and PriC have redundant roles. Thus *priA* and *dnaT* null mutants have extreme phenotypes whereas *priB* and *priC* null mutants have none. *dnaC809,820* is a *priC/rep*-independent suppressor that restores all known phenotypes of *priA* and *dnaT* null mutants (Sandler et al., 1999). During the lytic cycle of Mu growth, PriA restarts Mu replication without the involvement of homologous recombination ((Jones and Nakai, 1997; Jones and Nakai, 1999) and Fig. 3.5B). The data reported in Figs. 3.1-3.9 in this study show that PriA and DnaT are also required during the non-replicative event, along with a requirement for homologous recombination proteins.

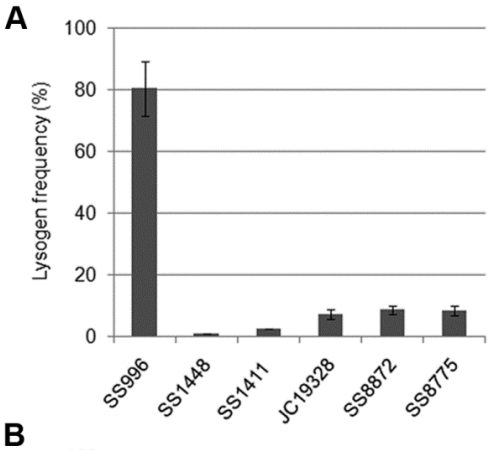
To confirm the phenotype of *priA*, *dnaT*, and the *rec* genes and to dissect the role of PriA further, we tested these and several different mutant alleles of these genes in a different strain background. The *priA*, *dnaT*, *recA*, *recB* (and *recBCD*) mutants all showed defects in Mu lysogen recovery in this strain background as well (Fig. 3.10A). *priA* and *dnaT* mutants show poor growth (Fig. 3.5A and (Lee and Kornberg, 1991; Nurse et al., 1991; McCool et al., 2004a)) and many cells in the population have high levels of SOS expression (McCool et al., 2004b). SOS genes are normally kept silent by the repressor LexA, and activated only when LexA is cleaved by RecA in response to

DNA damage (Kuzminov, 1999). SOS induction can be prevented by removing *recA* or by introducing a non-cleavable *lexA3* allele (Little et al., 1980). To test if SOS expression is responsible for low recovery of Mu lysogens, we tested *priA lexA3* and *priA recA* double mutants; both mutants remained defective (Fig. 3.10B). A *lexA3* mutant alone supported efficient recovery of Mu insertions, showing additionally that the SOS response is not required, but that the recombination function of RecA is needed. We note that *recA1*, a recombination-defective missense allele of *recA*, was not seen to affect recovery of Mu insertions in *Salmonella* (Hughes et al., 1987; Sonti et al., 1993). This allele can bind ssDNA *in vitro* (Lauder and Kowalczykowski, 1993), and perhaps has residual activity *in vivo* that allows it to function in Mu repair. We also note that several genes in the Keio collection were recently reported to be partially duplicated (Yamamoto et al., 2009). Of these, *priB* and *polA* are of interest to this study. These gene deletions as well as *priC* were therefore re-tested in the same strain background as the *priA* alleles. They were found to not affect Mu recovery (Fig. 3.10B).

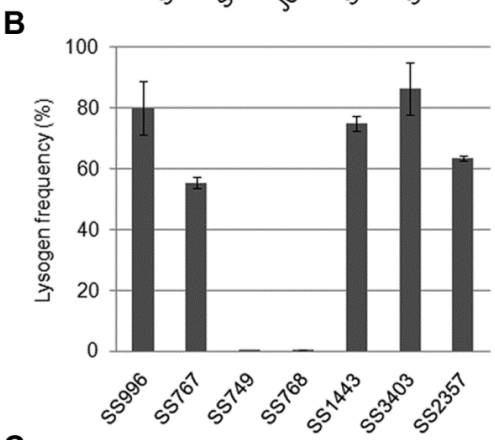
PriA has at least four types of activities: ATPase, helicase, the ability to load the replisome, and the ability to interact with other proteins. PriA300 (K230R) inactivates the ATPase and helicase activities, yet primosome assembly can occur both *in vivo* and *in vitro* (Zavitz and Marians, 1992; Sandler, 2005). PriA301 (C479Y) mutates a residue in the cysteine-rich region of PriA thought to be important for protein-protein interactions and helicase activity (Zavitz and Marians, 1993). Like *priA300, priA301* maintains wild-type growth and recombination proficiency (Sandler et al., 1996; Sandler et al., 2001). Lack of the helicase activity of PriA has been reported to impair Mu replication both *in vivo* and *in vitro* (Jones and Nakai, 1999). Using the helicase-defective strains *priA300* and *priA301*, we observed that the helicase and protein-protein interaction activities of PriA are largely dispensable (Fig. 3.10C), indicating that it is the primosome activity of

PriA that is essential for recovery of Mu insertions. This is further supported by the observation that combining *priA* and *dnaT* null mutations with *dnaC809,820* restores the ability of strains to recover lysogens (Fig. 3.10C). Both *in vivo* and *in vitro* experiments have suggested that mutant DnaC proteins suppress the absence of PriA/DnaT complex by bypassing its role in helping DnaC to load DnaB/PolIII directly onto a recombinational intermediate (Liu et al., 1999; Sandler et al., 1999).

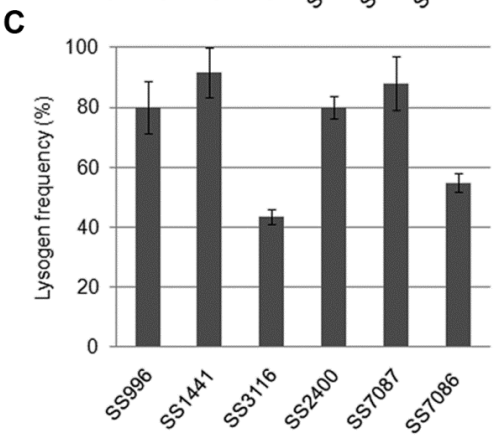
To confirm that all of these data point to a critical role for replication restart in repair of Mu insertions, we sequenced fifteen independent insertions which were recovered at a low frequency in the *priA* mutant (see Materials and Methods) (Fig. 3.11). Of these, five insertions had rearranged the Mu-host junctions in various ways, and their precise location could not be determined. Two insertions had symmetrical additions (at both ends) of a nucleotide not found in the wild type host, likely due to repair by an error-prone polymerase, and one of these strains had two copies of Mu. Eight insertions had normal Mu-host junctions. We note that the sequencing strategy included cloning of Cm^R Mu DNA fragments, favoring recovery of R end fragments that had not been deleted or rearranged, and therefore underestimating the fraction of incorrectly repaired insertions. Overall, these results show that in the absence of PriA, Mu insertions are repaired inefficiently and often incorrectly by alternate pathways. Thus, PriA is indeed required for normal repair of Mu insertions.



strain	genotype	lys freq (%)
SS996	*WT*	80.56
SS1448	*priA2*	1.24
SS1411	*dnaT822*	2.80
JC19328	*recA*	7.35
SS8872	*recB*	11.13
SS8775	*recBCD*	11.23



strain	genotype	lys freq (%)
SS996	*WT*	80.56
SS767	*lexA3*	55.49
SS749	*priA, recA*	0.04
SS768	*priA, lexA3*	0.38
SS1443	*priB*	74.96
SS3403	*priC*	86.33
SS2357	*polA*	63.37



strain	genotype	lys freq (%)
SS996	*WT*	80.56
SS1441	*priA300*	91.69
SS3116	*priA301*	43.48
SS2400	*dnaC809,820*	80.00
SS7807	*priA2, dnaC809,820*	88.07
SS7806	*dnaT822, dnaC809,820*	54.82

Figure 3.10 Behavior of various *priA*, *dnaT* and *rec* alleles in a different strain background.

Mu*Bam*1066ΔSE::Cm was used for infection of indicated strains to assess their role in recovery of Mu insertions. Other descriptions as in Figure 3.8.

No.	Orientation (5' → 3' from *ori*)	duplicated 5-bp at insertion site	Insertion site (bp)
1		-	-
2		-	-
3		-	-
4		•GCTGG repeat	-
5		- CCGAG/CCGAG	- 3527292 (l)
6		*CACCGC /*CACCGC	649424 (r)
7		TACCC/TACCC *GGCGGT/*GGCGGT	1407549 (r) 1656805 (l)
8		CAGTG/CAGTG	2047883 (l)
9		TCGGG/TCGGG	2359304 (l)
10		CTGGT/CTGGT	2740503 (l)
11		CCTTT/CCTTT	2814401 (l)
12		CCGGG/CCGGG	3204709 (l)
13		GGCAA/GGCAA	3775432 (l)
14		GCGGA/GCGGA	4261114 (r)
15		CCGGC/CCGGC	4541833 (r)

Figure 3.11 Sequence of Mu-host junctions at 15 insertions recovered in a *priA* mutant infected with Mu::Cm(*Bam*1066).

See Materials and Methods for sequencing details. Orientation refers to clockwise positions of Mu from *oriC*, which is at ~3.92 Mb; *ter* is at ~1.59 Mb. The numbers in the Insertion site column refer to nucleotides on the E. coli genome. Black bars, intact Mu with L an R ends indicated; Gray bars, truncated/duplicated Mu with only one end identified; Dotted lines, undetermined host DNA sequence; • a repeated sequence; * insertion of nucleotides not found in the host DNA; l, r, position of insertions in the left and right replicores, respectively.

DISCUSSION

Most transposable elements generate characteristic target site duplications flanking their insertion sites, as a result of staggered cuts in the target initially made by the transposase (Craig, 2002). For the large majority of known transposable elements whose transposition is not coupled to replication, it is not known how the single-stranded gaps left in the target after strand transfer are filled. For retroviruses and Line retroelements, double-strand break repair pathways (NHEJ, ATM, ATR) have been implicated (Gasior et al., 2006; Smith and Daniel, 2006; Suzuki et al., 2009; Yang et al., 2010). The present study finds that for Mu, in the non-replicative pathway, the gaps are repaired by the primary machinery for double-strand break repair in *E. coli* – the PriA primosome and homologous recombination proteins. This finding represents a radical change in thinking regarding Mu transposition in particular, and the transposition field in general. In the case of Mu, this is because one did not expect replicative functions to be involved in a transposition event that had been labeled 'non-replicative' in early studies. The original label was somewhat of a misnomer in that it described the replication status of Mu prior to integration (Liebart et al., 1982; Akroyd and Symonds, 1983; Harshey, 1984). However, discovery of flap DNA removal upon Mu integration (Au et al., 2006; Choi and Harshey, 2010) meant that a second round of transposition could not occur until the gapped strand transfer intermediate was repaired. This event is therefore clearly different from the target-primed replication that immediately follows strand transfer during the replicative pathway. Early experiments that established the non-replicative transposition pathway found limited replication near the ends shortly after integration of infecting Mu, consistent with the idea of gap-filling repair (Harshey, 1984). We note that

simple inserts generated using crude extracts and mini-Mu plasmids *in vitro* were also seen to have some replication associated with ends of Mu DNA, although it is not clear whether these simple insertion events are representative of the first integration event after Mu infection (Mizuuchi, 1984). The identification of replication restart proteins in the present study suggests a new pathway for gap repair. These findings should spur a re-examination of similar assumptions made for other transposons that transpose by non-replicative mechanisms.

Requirement for PriA in both replicative and non-replicative Mu transposition

There are three pathways for replication restart in *E. coli*: PriA–PriB, PriA–PriC, and PriC–Rep, which differ in their recognition of stalled forked structures (Gabbai and Marians, 2010). PriA plays an essential role in initiation of replication on the forked DNA intermediates generated during the lytic phase of Mu growth, using either the PriA–PriB or PriA–PriC pathway, in addition to the proteins that are required for *E. coli* chromosomal replication (Jones and Nakai, 1997; Jones and Nakai, 1999; Jones and Nakai, 2000; North and Nakai, 2005). During Mu transposition, the transition from strand transfer to DNA replication can be divided into a number of discrete steps (Nakai et al., 2001; Chaconas and Harshey, 2002). MuA initially remains tightly bound to the Mu fork as a multi-subunit complex called transpososome. In a highly choreographed series of steps, host proteins dislodge this transpososome and assemble a replisome. In the first step of this transition, ClpX alters MuA subunit interactions to weaken interaction of the transpososome with DNA (Levchenko et al., 1995; Kruklitis et al., 1996; Abdelhakim et al., 2010). Next, as yet unidentified cellular factors called Mu Replication Factor α2 (MRFα2) displace the transpososome and exchange it with the translation initiation

factor IF2-2 to produce a pre-replisome (North et al., 2007). Finally, the helicase activity of PriA is required to displace IF2-2, remodeling the template to permit replisome assembly, which includes DnaT, DnaB, DnaC and the DNA polymerase III holoenzyme (Nakai et al., 2001). PriA has distinct replisome assembly and 3' to 5' helicase activities (Gabbai and Marians, 2010). Helicase-defective PriA supports little or no Mu replication *in vitro*, and shows a partial defect in Mu replication *in vivo* (Jones and Nakai, 1999). These data indicate that PriA's replisome assembly activity is essential for initiation of Mu DNA replication and that the helicase activity also promotes this process. PriA is thought to bind to the lagging strand template at the fork and unwind it in a 3' to 5' direction, promoting loading of DnaB, thus coupling its replisome assembly and helicase activities.

The surprising requirement of PriA and DnaT in the non-replicative pathway of Mu transposition as reported in this study, suggests strongly that the 5 bp gaps generated upon Mu insertion are repaired by the replication restart machinery. This shared requirement for the PriA primosome in both pathways might imply that the PriA loading steps after strand transfer are similar in both. What apparently distinguishes the two pathways is non-requirement of the helicase activity of PriA, and requirement for homologous recombination proteins. We discuss two alternate models for recombinational gap repair below.

Models for Recombinational Repair

Nicks and gaps in DNA are normally repaired when their encounter with a traveling replication fork converts them into a double strand break, collapsing the fork (Kuzminov, 1999). The broken end serves as an entry point for RecBCD, generating

single strands for RecA binding, followed by invasion of the intact sister chromosome, thus reconstituting a forked structure for restarting replication via the PriA primosome (McGlynn and Lloyd, 2002; Michel et al., 2004). In such a scenario for Mu repair, an *oriC*-initated fork will cause a double strand break when, arriving at the site of a Mu insertion, it encounters the flanking gap (Fig. 3.12A). The double-strand break will be on the chromosomal DNA flanking the Mu insertion, which is expected to be processed by RecBCD, followed by restoration of the fork by recombination, and restart of replication by the primosome. Two considerations make this scenario unappealing. First, Mu does not insert near replication forks (Nakai and Taylor, 1985), so the unrepaired intermediate would be potentially vulnerable to degradation while it waits for the *oriC*-initiated fork to arrive. Second, the passing fork would encounter only one of the two gaps at each Mu end that need repair, so the entire Mu would have to be replicated, generating a second double strand break at the distal Mu end, reiterating RecA-mediated invasion and primosome assembly before repair of that gap can be completed. A parsimonious alternative model takes advantage of the PriA replisome already present at the forked strand transfer joints at both Mu ends, recruited there in the normal course of transpososome disassembly (see Fig. 1.6). In this model, the initial steps of PriA recruitment and replication are common to both the repair and replication pathways (Fig. 3.12B). The pathways differ in the flap cleavage step, which ensues concomitant with replication restart, leaving double-strand breaks on the Mu lagging strand. These breaks allow RecBCD entry, creating single-stranded 5’ Mu ends on which RecA polymerizes (Register and Griffith, 1985). Although 3’ end strand invasion is generally preferred with purified RecA, 5’ ends can be used for strand exchange *in vitro* (Bork et al., 2001), and *in vivo* recombination data also fit models that invoke 5’ strand invasion (Gumbiner-Russo and Rosenberg, 2007). The Holliday junction so created can then be resolved by Ruv

proteins or endonucleases. This model reverses the steps normally associated with recombinational repair, with replication preceding recombination. According to this model, there will be limited replication near the two Mu ends in this largely non-replicative event.

What signals flap cleavage in one pathway and not in the other? We speculate that the MuN protein, which normally protects the ends of infecting Mu DNA from degradation, dissociates from the ends, perhaps upon interaction with the transpososome assembled on the strand transfer complex. This allows RecBC to enter and peel away the 3’ strand of the flap, engaging and activating MuA_{Nuc} on the 5’ strand.

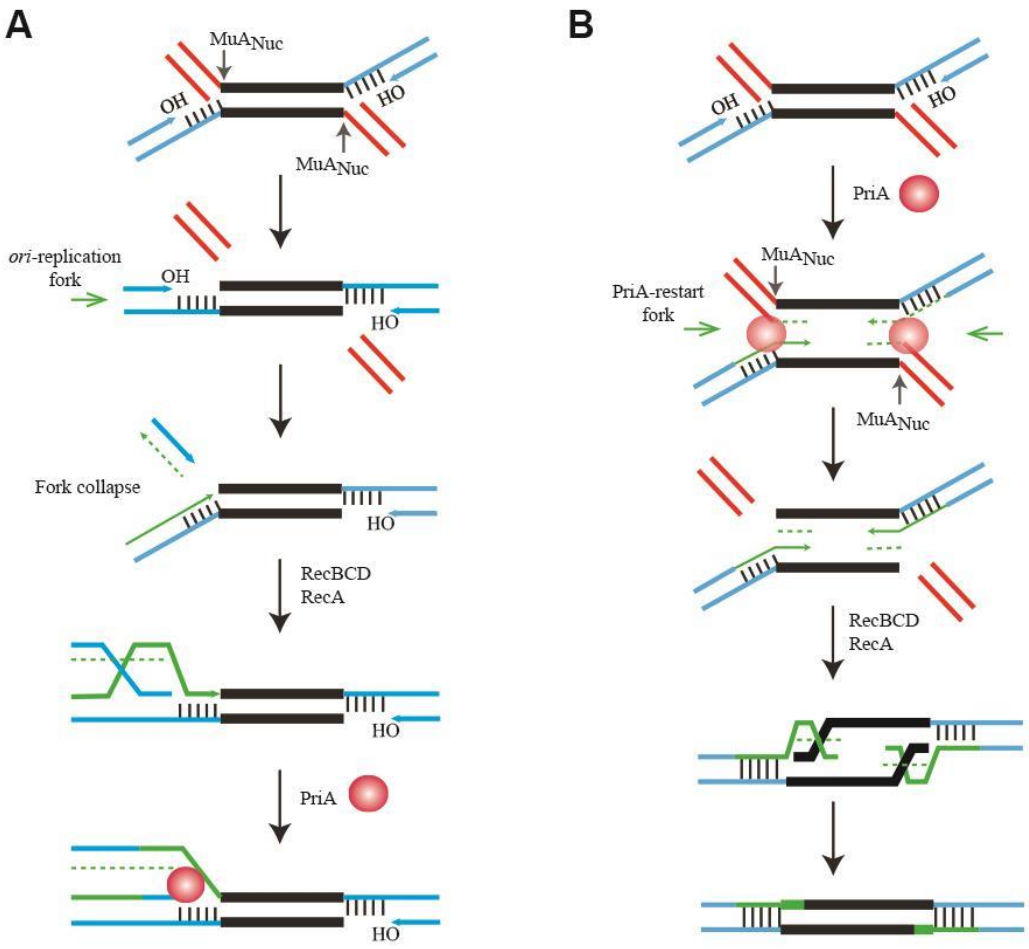


Figure 3.12 Models for recombinational repair of Mu insertions in the non-replicative pathway.

Both models rely on repair of double strand breaks by homologous recombination and replication restart proteins, but differ in the location of the break and the order of the recombination/ restart-replication events that follow. In (**A**), the break is on the chromosomal DNA flanking the Mu insertion. Here, homologous recombination is followed by restart replication. In (**B**), the break is on the Mu lagging strand. Here, restart replication precedes homologous recombination. See text for details.

SUMMARY

This is the first report of specific host processes involved in repair of transposon insertions in bacteria. We find that the PriA primosome and homologous recombination proteins, which are essential for repair of double-strand breaks in *E. coli*, play a critical role in the repair of Mu insertions. We favor a model for recombinational repair of Mu insertions where PriA restart of Mu replication is followed by RecA-mediated resolution of double-strand breaks on the Mu lagging strands created by the flap endonuclease activity of the transposase. Given that the predominant route taken by Mu upon infection is to enter lytic growth, it is plausible that Mu first co-opted the PriA system for replication, and later used it for repair. It will be interesting to see whether other transposons use these same processes for repair of their insertions.

Chapter 4. Repair of Mu insertions begins only when the *E. coli* replisome collides with the transpososome, leaving double strand breaks in its wake

ABSTRACT

The movement of transposable elements creates short flanking gaps in the target DNA, which must be repaired to complete transposition. It has been assumed that these gaps are filled by gap-filling polymerases. Taking advantage of the high efficiency of phage Mu transposition and of a unique feature of its structure, we have discovered that the Pol III replisome is essential for repair. Unrepaired Mu insertions persist indefinitely when the replication fork is arrested. Release of the fork generates double strand breaks (DSBs) proximal to Mu. These findings suggest that interaction of the replisome with the transpososome is required to expose the gaps for repair, and that the replisome stalled at the DSB is specifically exploited by the transpososome to coordinate repair of both flanking gaps without replicating the intervening Mu DNA. The findings are of broad significance because DNA transposons, retrotransposons and retroviral elements (e.g. HIV-1) share the common problem of gap repair following transposition.

INTRODUCTION

Genomes of virtually all organisms harbor transposable elements (TEs) whose past as well as present activity continues to shape genome structure, function and evolution (Huang et al., 2012). Active human TEs have been estimated to generate about one new insertion per 10-100 human births (Kazazian, 1999). In single individuals, a significant number of *de novo* insertions influence a range of phenotypes, both life enhancing (creating somatic heterogeneity in the brain; (Singer et al., 2010; Perrat et al., 2013)), and life threatening (primarily cancer-causing; (Kazazian, 2004; Mills et al., 2007; Chenais, 2013)). Understanding the mechanism and regulation of these events is important for controlling their incidence.

The cutting and joining reactions of transposition that link the transposon to the target are well studied (Craig, 2002; Chandler and Craig, 2015). The majority of DNA transposons, including retroviruses and retroviral-like transposons, transpose by a non-replicative mechanism i.e. without duplicating themselves in the process. These reactions leave short gaps in the target DNA on either side of the transposon. Transposition is not complete until these gaps are repaired. Yet this essential step is not as yet deciphered. Because a majority of these transposons are not duplicated, it is assumed that gap-filling polymerases fill the gaps, but conclusive evidence for a specific polymerase is lacking (Sasakawa et al., 1981; Syvanen et al., 1982; Yoder and Bushman, 2000). In this study we have used transposable phage Mu to investigate gap repair i*n vivo*, because of unique features of Mu that makes the analysis possible.

As a temperate transposable phage, Mu uses transposition to integrate into the *E. coli* host chromosome to generate a prophage during the lysogenic phase, and to amplify

its genome over a hundred-fold during the lytic phase (Symonds et al., 1987). During both phases, the chemical steps of transposition are the same: single-stranded DNA cleavages at Mu ends followed by strand transfer (ST) of the cleaved ends to phosphodiester bonds spaced 5 bp apart on the target (Mizuuchi, 1992; Chaconas and Harshey, 2002). However, the structure of the Mu donor is different during the two phases, and has a bearing on how the ST intermediate is resolved (Fig. 4.1 and 4.2A) (Harshey, 2015). During the lytic phase, Mu transposes from one site to another on the *E. coli* chromosome while remaining part of the covalently closed chromosome (Fig. 4.2B, 60 min). Here, the ST intermediate is resolved by target-primed replication by the Restart primosome, which fills the flanking target gaps while replicating across Mu (Fig. 4.1) (Nakai et al., 2001). During the lysogenic phase, the infecting Mu genome is linear, is linked to several hundred base pairs of non-Mu flanking DNA (FD), and is non-covalently closed by the phage N protein (Fig. 4.2A) (Harshey and Bukhari, 1983; Puspurs et al., 1983; Gloor and Chaconas, 1986). Here, the ST intermediate is resolved without replication (Liebart et al., 1982; Akroyd and Symonds, 1983; Chaconas et al., 1983; Harshey, 1984), during which the FD is degraded concomitant with gap repair (Au et al., 2006). Using a convenient assay for monitoring FD degradation *in vivo*, we have learned that the first event in repair is removal of the FD by the RecBCD exonuclease (Choi et al., 2014a), whose entry past the N-protein block is controlled by the transpososome and facilitated by ClpX (Choi and Harshey, 2010; Choi et al., 2014b). *In vitro* experiments reveal that RecBCD action is required for stimulating endonucleolytic cleavage within the transpososome-protected DNA, leaving 4-nt flanks outside both Mu ends (Fig. 4.2A). This structure is likely the substrate for gap repair by host enzymes. The infection phase of non-replicative Mu transposition is an ideal system to investigate the gap repair process not only because of its high efficiency, where every infecting Mu

genome integrates into the *E. coli* chromosome within 10 minutes, but also because we can track repair events using the FD degradation assay. We know that when integration of infecting Mu is blocked, the unintegrated N-linked Mu genome is indefinitely stable (Harshey and Bukhari, 1983; Puspurs et al., 1983; Gloor and Chaconas, 1986). Thus, degradation of the FD is timed to coincide with some event that follows integration. We demonstrate in this study that this event is arrival of the *E. coli* replication fork.

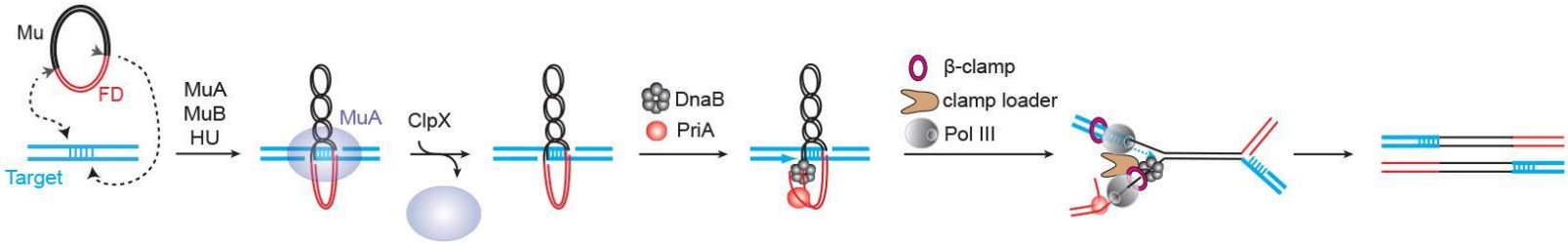


Figure 4.1 Known steps in the replicative pathway of Mu transposition.

Transposition is depicted as an intermolecular event, as studied *in vitro* on plasmid substrates (Mizuuchi, 1992; Nakai et al., 2001). This pathway is used to amplify the Mu genome during lytic growth *in vivo*, where these events are intramolecular. Transposition begins when MuA transposase introduces single-stranded cleavages at the 3' ends Mu, and transfers the resulting 3'OHs to phosphodiester bonds spaced 5 bp apart in the target DNA, assisted by MuB protein and host HU protein. The transpososome (purple ball) is first destabilized by the molecular chaperone ClpX, and exchanged with a series of host factors, culminating in PriA primosome-assisted loading of DnaB helicase and the Pol III holoenzyme on one Mu end, followed by target-primed replication across Mu (Jones and Nakai, 1999; Nakai et al., 2001; North et al., 2007; Abdelhakim et al., 2010).

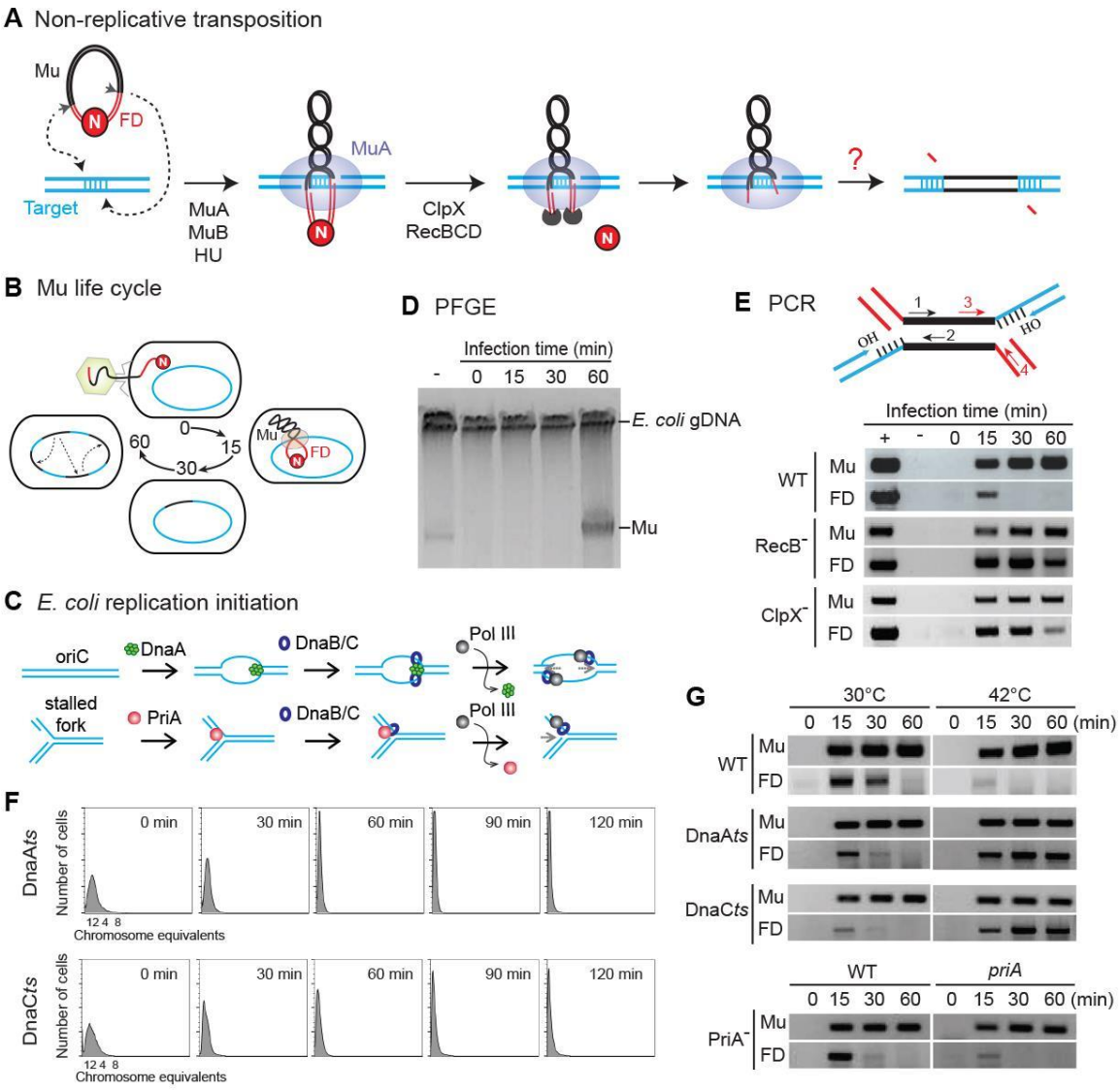


Figure 4.2 Repair of Mu insertions is prevented if the *oriC* replication fork is blocked.

(**A**) Known steps in the non-replicative (repair) pathway of Mu transposition. This pathway is used during integration of infecting Mu. The infecting genome is linear, and attached at both ends to long flanking DNA (FD) protected by Mu N protein. This DNA is variable in length (60 - 150 bp at the L end and 0.5 – 3 Kbp at the R end). MuN circularizes the DNA non-covalently, and protects it from nucleases. MuA catalyzes

-Figure 4.2 *legend continued*-

cleavage and strand transfer (integration) of Mu into the *E. coli* genome, assisted by MuB protein and host HU protein. The N protein is removed only after integration by an unknown mechanism assisted by the transpososome (purple ball), and the FD is degraded by RecBCD. Degradation is slowed in the absence of ClpX. *In vitro*, the final product of RecBCD degradation is +4 nucleotides. This strand transfer intermediate with short flanks is likely the substrate for the final steps in repair, where the 5 bp target gaps are filled to generate a simple insertion. (**B**) Mu life cycle. After infection, Mu integrates into the *E. coli* genome, the FD is degraded, the Mu insertion is repaired, and Mu enters the lytic cycle. The approximate time (0-60 min) of these events is indicated. (**C**) Schematic of known mechanisms for replication initiation at *oriC* and during Restart of stalled forks in *E. coli*. (**D**) Preparation for the FD detection assay. At various times after infection, genomic DNA was subjected to pulse-field agarose gel electrophoresis (PFGE) to separate integrated Mu from free Mu, and the gDNA was excised for analysis by PCR. The (–) lane is an uninfected control where Mu DNA was added to the gDNA prior to electrophoresis, to assess contamination of the excised gDNA band with free Mu. Mu-length DNA at 60 min reflects packaged virions. (**E**) PCR assay for FD detection. Strains were infected with Mu at 37℃ for the indicated times, and the isolated gDNA was tested by PCR to detect Mu or FD (*lacZ*) sequences using appropriate primers; *lacZ* sequences linked to infecting Mu are not found in the host (Au et al., 2006). + is Mu virion DNA and – is gDNA from the uninfected control lane in panel D. WT (BW25113); RecB$^-$ (JW2788); ClpX$^-$ (JW0428). (**F**) Monitoring replication arrest in DnaA*ts* (SS1424) and DnaC*ts* (SS1021) mutants by measuring chromosome equivalents. The mutants were held at 42℃ for indicated times, without shaking, and fixed in ethanol before staining with the fluorescent DNA stain SYTOX Green. Stained cells were analyzed by flow cytometry as

-Figure 4.2 *legend continued*-

described under Materials and Methods. By 60 min, there are no new rounds of replication in either mutant, as judged by the shift in the initial DNA content of ~2-4 chromosome equivalents to 1 chromosome equivalent. (**G**) FD removal depends on the native replication fork. DnaA*ts* and DnaC*ts* mutants were infected with Mu at both 30℃ and 42℃. The latter infections were carried out after replication arrest for 90 min. In the WT (MG1655) infection at 42℃, FD is removed faster than at the lower temperatures. The PriA mutant (SS1448) and its WT parent (SS996) were infected at 30℃. Strains are listed and described in Table 2.1.

RESULTS

Repair of Mu insertions depends on the replication fork

The cutting and joining reactions of Mu transposition do not generate DSBs (Mizuuchi, 1992), yet the DSB repair pathway of *E. coli* is required to recover Mu insertions after infection (Jang et al., 2012). This suggests that one of the Mu target gaps might be converted to a DSB during repair of the Mu insertion, perhaps by a replication fork (technically, this would create a double strand end or DSE, but we will refer to it as DSB for convenience). We had envisioned two alternative scenarios for generating the break - either arrival of the native fork at the Mu insertion site, or start of replication on the FD-linked Mu insertion (Jang et al., 2012). In *E. coli*, native replication initiates at a unique origin *oriC* (Kornberg and Baker, 1992), where the initiator protein DnaA first recruits the DnaB helicase from the DnaB-DnaC complex, followed by loading of the Pol III holoenzyme (Fig. 4.2C, top). Bidirectional replication forks proceed from *oriC* to the terminus *ter*, but can stall if they encounter nicks, gaps or other forms of DNA damage (Michel et al., 2007). Collapsed forks are reassembled by several Restart pathways (Gabbai and Marians, 2010), the two main pathways requiring PriA and DnaT. The function of Restart proteins is similar to DnaA in recruiting DnaB for Pol III assembly (Fig. 4.2C, bottom). Mu uses the PriA Restart primosome to initiate replication specifically on Mu DNA during lytic growth (Fig. 4.1) (Nakai et al., 2001).

To determine the role of the replication enzymes in the repair of Mu-linked FD, we selectively blocked either the *oriC* fork by using temperature sensitive (*ts*) mutants of DnaA and DnaC, or the Restart fork by using a PriA mutant, and monitored FD repair. At various times after Mu infection, *E. coli* genomic DNA (gDNA) was subjected to pulse-

field agarose gel electrophoresis to separate integrated Mu from free Mu (Fig. 4.2D). The gDNA band was excised from the gel, and a PCR assay was used to detect Mu (primers 1, 2) and FD sequences (primers 3, 4) (Fig. 4.2E). In a wild type host infection, both Mu and FD are found integrated in the gDNA by 15 min, but the FD disappears soon after (Fig. 4.2E; WT). The FD is not degraded in a RecB mutant (Fig. 4.2E; RecB$^-$), and degradation is delayed in a ClpX mutant, as reported earlier and (Fig. 4.2E; ClpX$^-$) (Au et al., 2006; Choi and Harshey, 2010; Choi et al., 2014b). A different integration assay showed additionally that the FD was processed to a short length *in vivo* (Fig. 4.3), similar to that seen *in vitro* with RecBCD (Fig. 4.2A) (Choi et al., 2014a).

When *oriC* replication is arrested in DnaA*ts* and DnaC*ts* mutants at 42℃, the *ts* mutants finish ongoing rounds of replication but do not re-initiate new ones (Wechsler and Gross, 1971), an observation we reconfirmed using FACS analysis for measuring chromosome ploidy, which shifts from two chromosome equivalents to one within ~ 60 min (Fig. 4.2F). After replication arrest, the cells were infected with Mu. Both mutants supported Mu integration at 42℃ (Fig. 4.2G; see Mu panels in the Dna*ts* mutants), consistent with earlier data showing that Mu integration is independent of replication forks (Nakai and Taylor, 1985). However, in the absence of the *oriC* fork, the FD was not removed (Fig. 4.2G, compare FD panels at 30°C vs 42° in the *ts* mutants). The FD was degraded normally in the absence of PriA. These results were confirmed using an alternate assay, where the short product of FD degradation was not detected in replication-arrested cells (Fig. 4.4). We conclude that the native replication fork is required to begin repair of Mu insertions, and that PriA-dependent Mu replication is not involved.

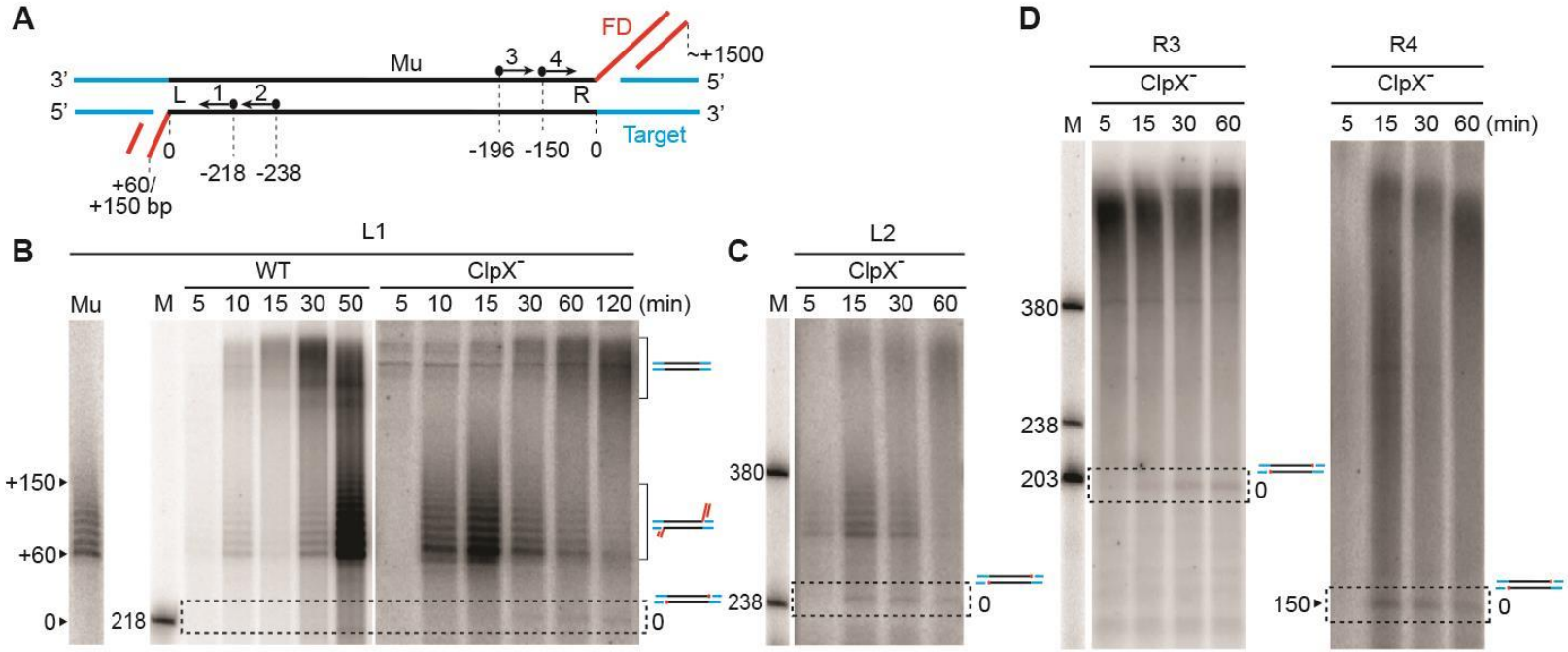


Figure 4.3 Delayed FD removal in a ClpX mutant allows detection of a specific 5' cleavage product near the Mu-FD junction.

(**A**) Schematic showing position of primers annealing to the L and R ends of Mu, and the variable lengths of FD (red), which range from 60-150 bp on the L end and 0.5- 3Kbp on the R end (Symonds et al., 1987). Radiolabeled primers were used to perform a run-off IPE assay (iterative polymerase extension) shown in B. This assay monitors FD in all the infecting genomes, not just in select *lacZ* FD sequences followed in Figure 4.2. (**B**) WT (BW25113) and its ClpX mutant derivative (JW0428) were infected with Mu at 37℃. The ClpX mutant could be followed for a longer time because Mu does not enter the lytic

-Figure 4.3 *legend continued*-

cycle in this mutant. FD removal is delayed in a ClpX mutant (Fig. 4.2E and (Choi and Harshey, 2010)). We used this delay to observe additional *in vivo* repair events closer to the Mu-FD junction, because *in vitro* experiments show that the final FD degradation product is +4 (Choi et al., 2014a). Total genomic DNA from WT and ClpX$^-$ strains was isolated at indicated times, subjected to IPE with radiolabeled L1 primer, and resolved on polyacrylamide gels (see Materials and Methods). The L-end primers will detect a ladder of bands ranging from +60 to + 150 bp corresponding to the variable FD lengths at the L end (George and Bukhari, 1981). In a WT infection, this ladder was most prominent at 10 min, diminished at 15 min, and reappeared starting at 30 min due to onset of phage maturation (see also Fig. 4.2D, 60 min). In a ClpX mutant, where Mu does not replicate, the ladder was most prominent at 15 min, diminishing gradually over 60 min. Around 15 min, a smaller product indicative of a shortened FD was observed (marked 0). The 218 nucleotide size marker corresponds to primer extension up to the left Mu-FD junction marked 0 in A. We expect the actual product to be +4 nucleotides longer. Completion of repair would yield the higher molecular weight products. (**C**) When the primer annealing site was moved 20 bp upstream (L2), the length of this short product increased accordingly. The DNA is from the ClpX$^-$ infection. (**D**) At the R end, the variable FD is too large to be resolved (0.5 - 3 Kbp; see A). However, two sets of R-end primers (R3 and R4; see A) identified a product corresponding to cleavage close to the right Mu-FD junction similar to that at the L end. The amount of the product remained constant suggesting that it is repaired soon after it is generated. As observed *in vitro* (Choi et al., 2014a), the short product was not observed in the absence of bulk FD removal (see Fig. 4.4). The data show that the FD is cleaved close to the Mu-host junction in the short product, but do not allow us to determine its exact position.

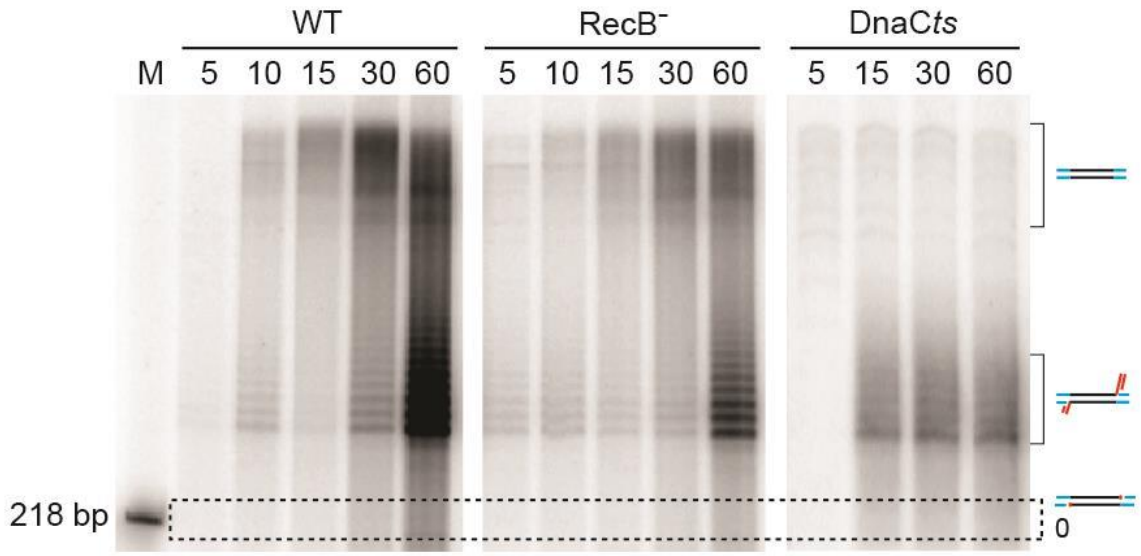


Figure 4.4 The 5’ FD cleavage close to the Mu junction is not generated in the absence of bulk FD degradation.

As described in Figure 4.3, total genomic DNA was subjected to IPE with radiolabeled L1 primer, and the products resolved on polyacrylamide gels. FD is not degraded in a RecB mutant (Fig. 4.2E and (Choi et al., 2014a)), or when the replication fork is arrested in the DnaC*ts* mutant (Fig. 4.2G). In the RecB mutant, the FD ladder does not disappear but gets prominent at 60 min due to phage maturation. The WT panel from Figure 4.3B is shown for comparison. We infer that in WT, the expected short (0) FD product must be repaired too quickly to be detected, because it is detected in a ClpX mutant where FD degradation is delayed (Fig. 4.3). RecB$^-$ (JW2788); DnaC*ts* (SS1021).

The entire Pol III replisome machinery is necessary for FD removal: gap-filling polymerases are not required

To determine if a specific component of the replisome is required for FD removal, *ts* mutants in all available components – helicase DnaB, β-clamp DnaN, β-clamp loader DnaX, and Pol III α subunit DnaE – were tested (Fig. 4.5A). Replication ceases immediately at 42°C in these replication elongation mutants (Saluja and Godson, 1995). Mu integration was apparently normal when replication was arrested (Fig. 4.5B, 42°C and Fig. 4.6); however, FD degradation was blocked in all the strains only at 42°C.

E. coli has four other polymerases that assist in gap repair either during normal replication (Pol I), or during repair of DNA damage (Pol II, IV, V) (Fijalkowska et al., 2012). Both Mu integration and FD degradation were similar to wild type in the single Pol I mutant as well as in the triple damage-inducible polymerase mutants (Fig. 4.5C). Thus, none of the gap-filling polymerases were required for Mu repair. These data do not address whether Mu integration might trigger an SOS response, the damage-inducible polymerases serving as a back-up repair mechanism. To test this, we measured SOS induction by monitoring GFP fluorescence expressed from the promoter of the SOS-induced gene *sulA* with or without Mu infection in a wild type strain, or its *lexA3* and *lexA71* derivatives (Fig. 4.5D) (McCool et al., 2004b); *lexA3* is defective for SOS induction, and serves as negative control, while *lexA71* is constitutively induced. The data show that Mu infection does not trigger an SOS response (Fig. 4.5D). We conclude that the Pol III holoenzyme, and not gap-filling polymerases, carries out Mu repair.

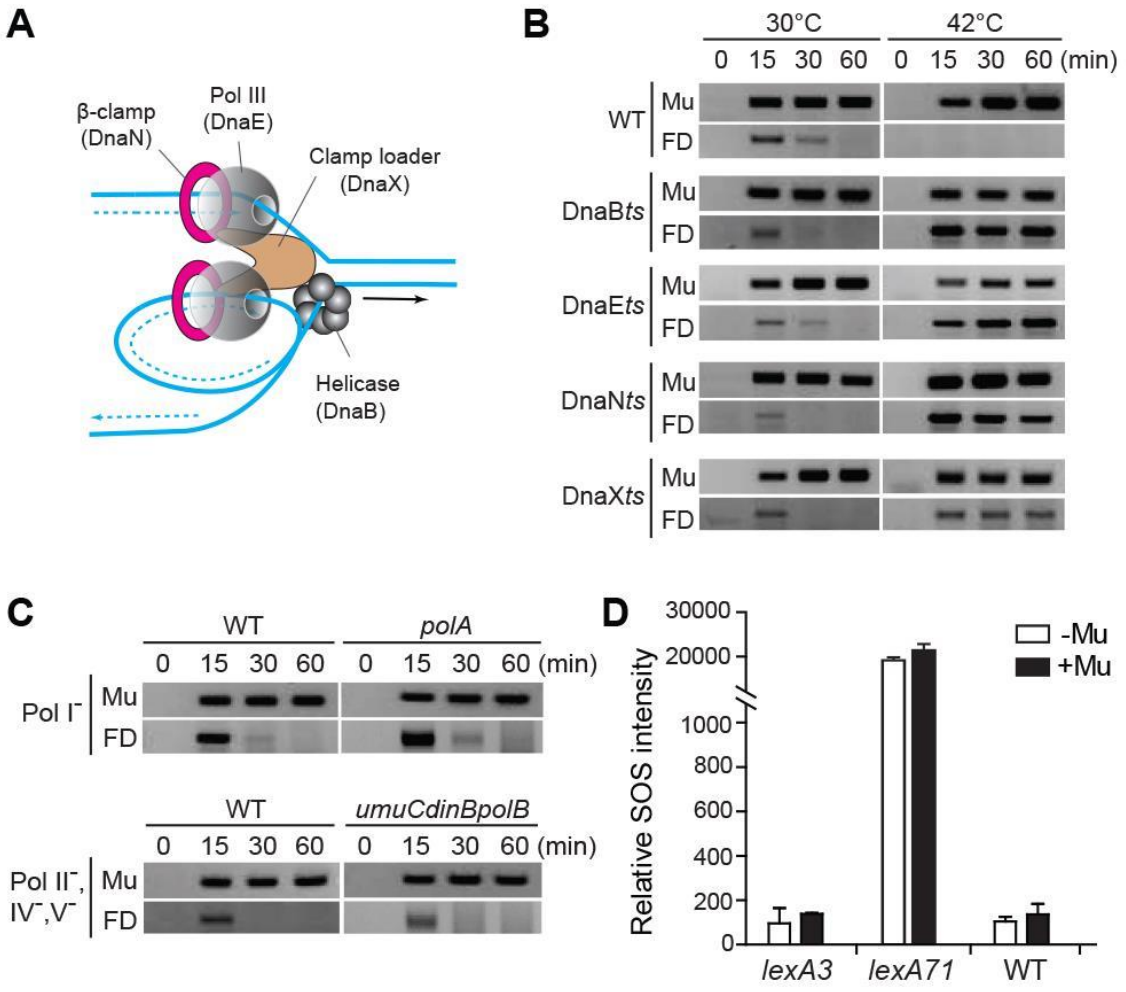


Figure 4.5 The entire Pol III replisome machinery is necessary for FD removal: gap-filling polymerases are not required.

(**A**) Replisome components tested in this experiment. Newly synthesized DNA is depicted by dotted lines. (**B**) Prior to Mu infection, replication was arrested in the *ts* mutants by incubating cells at 42°C for 30 min. Mu and FD sequences were monitored as described in Figure 4.2D and E. WT (SS996); DnaB*ts* (SS6699); DnaE*ts* (SS6239); DnaN*ts* (SS6700); DnaX*ts* (SS6698). (**C**) The *polA* (Pol I) mutant (SS2357) and its WT parent were infected with Mu at 30°C, while the triple *umuC dinB polB* (Pol II, IV,V) mutant (SS7346) and its WT parent were infected at 37°C, prior to analysis of Mu and

-Figure 4.5 *legend continued*-

FD. (**D**) SOS induction with or without Mu infection was monitored by expression of GFP under the control of the *sulA* promoter (*PsulA-gfp*) as described under Materials and Methods (McCool et al., 2004b). Mu infections were carried out in *lexA3* (SS4294) and *lexA71* (SS4610) derivatives as well as in their WT parent, all carrying a chromosomal *sulA-gfp* fusion. *lexA3* is defective for SOS induction, while *lexA71* is constitutively induced.

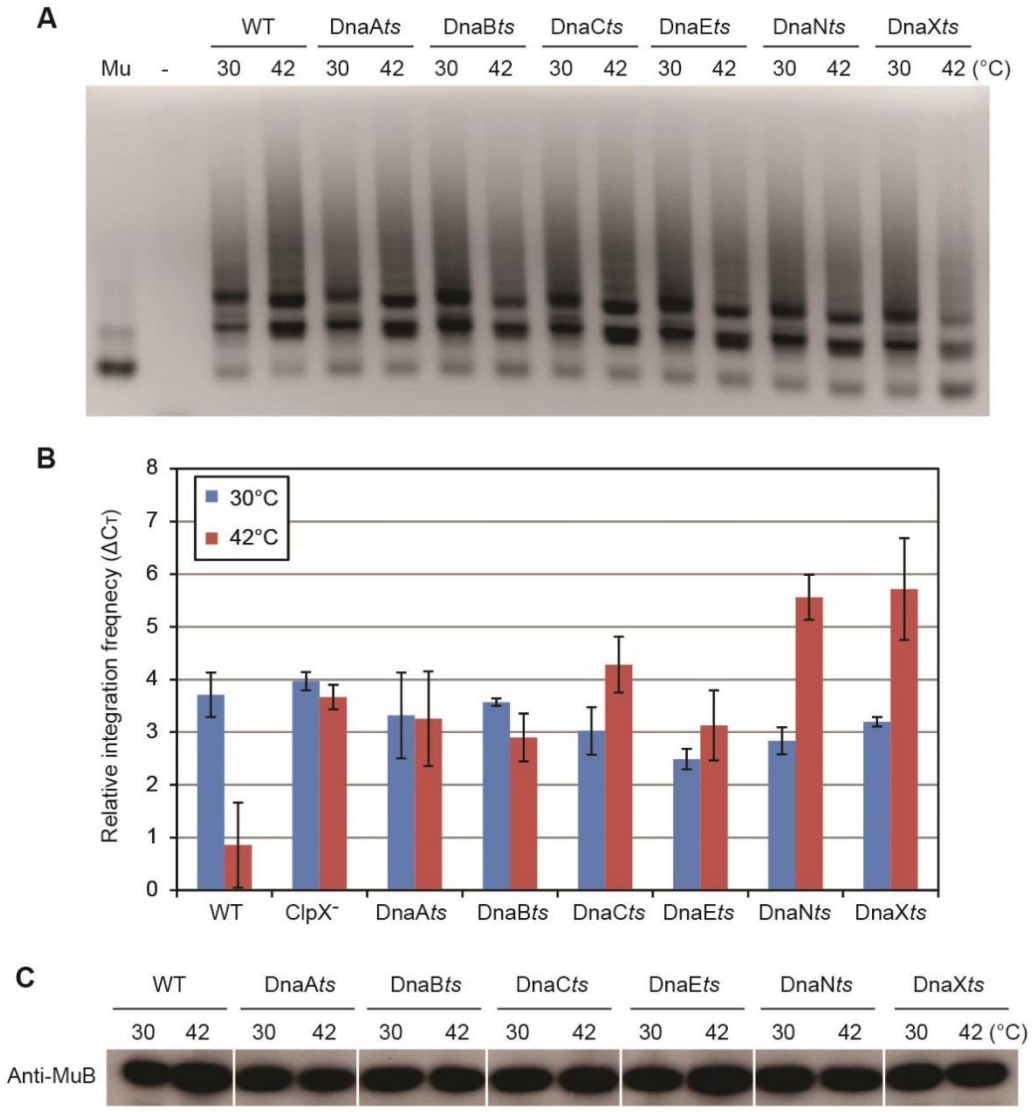


Figure 4.6 Mu integration and gene expression are normal in a majority of the Dna*ts* mutants.

(**A**) Standard PCR assay monitoring Mu integration efficiency in the *purH* gene 30 min after infection of the indicated mutants at the two temperatures; replication in the *ts* mutants was blocked as described in Materials and Methods (Jang et al., 2012). Primers were chosen to amplify covalent junctions between the left end of Mu DNA and an arbitrarily chosen target gene *purH*. PCR products of different lengths are expected since

-Figure 4.6 *legend continued*-

Mu integration is essentially random. A control reaction with Mu virion DNA alone is also shown. The reaction products were electrophoresed on 1% agarose gels and visualized by ethidium bromide staining. (**B**) Quantitation of Mu DNA integration by real-time PCR analysis. C_T is the fractional cycle number at the beginning of the exponential reaction phase where the SYBR Green fluorescence passes a threshold (T) at which the signal is first detected. 16S rRNA was used as a control for a single-copy chromosomal gene. The relative integration frequency is calculated as ΔC_T=Mu C_T – 16S rRNA C_T. The data are an average of three biological and technical repeats. C_T values are inversely proportional to the amount of DNA in the sample. Mu integration appears to be the highest at 42°C in the WT strain because Mu has begun to replicate. The ClpX mutant is a control for integration levels without Mu replication. Integration in the *ts* strains DnaN (β-clamp) and DnaX (clamp loader) is 50% reduced at the non-permissive temperature. (**C**) Western blot of Mu*B* gene expression. To ascertain that expression of Mu transposition functions was not altered in the various Dna*ts* host strains, levels of the essential transposition protein MuB were monitored after 30 min of infection as described under Materials and Methods. All lanes contain lysates derived from equal numbers of cells. WT (SS996). See Table 2.1 for other strain description.

Permitting blocked forks to resume replication, resumes Mu repair

We interpret the above data to mean that interaction of the *E. coli* replication fork with the Mu transpososome is required to allow entry of RecBCD into linear FD ends that are normally protected by the phage N protein (Fig. 4.2A). To test this interpretation, Mu was first allowed to integrate in replication-arrested cells, followed by release of the fork stalled either at the *oriC* site (DnaC*ts* mutant) or at any elongation site (DnaE*ts* mutant) on the chromosome, by a timed shift-down to 30℃. The DnaA*ts* mutant was not used in these experiments because Mu can enter the lytic mode of growth in this mutant (Toussaint and Faelen, 1974; McBeth and Taylor, 1982), during which the Restart fork would replicate Mu, bypassing FD repair, as also seen in RecBCD and other mutants which permit Mu lytic growth without FD repair ((Choi and Harshey, 2010; Choi et al., 2014a); see also Fig. 4.4)).

In the absence of native DNA replication, the FD remained stable at 42℃ when monitored across a 2 hr window in both DnaC and DnaE mutants (Fig. 4.7A, top panel; 120 min time-point not shown). When replication-blocked cells harboring integrated but unrepaired Mu in their genomes were shifted down to 30℃ after 15, 30 and 60 min at the non-permissive temperature, the FD was degraded in both mutants within 15-30 min after the shift-down (Fig. 4.7A, bottom 3 panels).

For *E. coli* growing in a rich medium, it takes ~40 min for the bi-directional fork to travel from *oriC* to *ter* (Fig. 4.7B) (Cooper and Helmstetter, 1968). Since Mu integration events occur all over the *E. coli* chromosome (Manna et al., 2004; Ge et al., 2011), on average, the fork would encounter a Mu insertion in approximately half that time, consistent with the average observed time for FD degradation. We conclude that Mu waits for the replication fork to begin repair of its insertions.

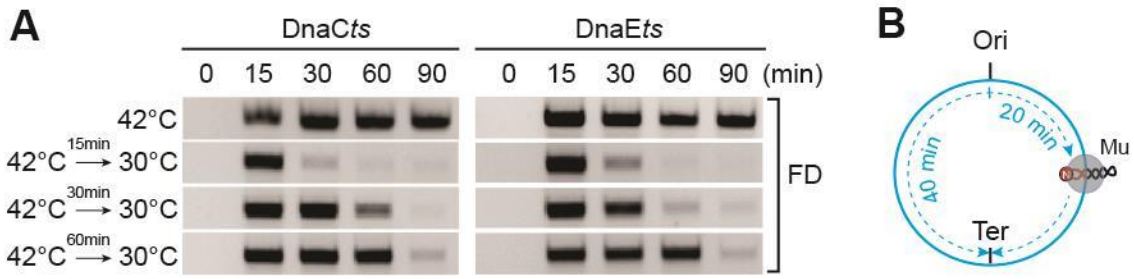


Figure 4.7 Permitting blocked forks to resume replication, resumes Mu repair.

(**A**) The indicated *ts* mutants were infected with Mu after fork inactivation at 42℃ for 30 min (DnaE*ts*) or 90 min (DnaC*ts*). Following infection (0 min), cells held at 42℃ (top row) were shifted down to 30℃ (15 - 60 min, rows 2-4) to release the replication block, and monitored for presence of the FD. (**B**) Schematic depicting the time taken by the bidirectional fork originating at *oriC* to reach *ter*. Mu integrates all over the chromosome. On average, the fork would encounter Mu in ~20 min.

FD is degraded by replication forks initiating at *oriC*-independent sites

Does a fork have to initiate at *oriC* to participate in FD repair? To address this question we used an RNase HI mutant (*rnhA*), where *oriC*-independent initiation can occur from RNA-DNA hybrids (R-loops) at multiple ectopic *oriK* sites using the PriA pathway (the R-loops are normally removed by RNase HI; (Kogoma, 1997)). The mutant strain grows poorly because of fork collisions associated with unregulated replication, and because of an SOS-constitutive phenotype brought on DNA breaks (Fig. 4.8A, B) (Maduike et al., 2014). The DnaA*ts* mutant was moved into this strain to additionally block *oriC* replication; the double mutant grows even more slowly (Fig. 4.8A). Nonetheless, both mutants still supported Mu integration at 30℃ as well as at 42℃, and although slightly delayed, the FD was clearly degraded at both temperatures (Fig. 4.8C). We conclude that the FD is degraded even by forks that do not initiate at *oriC*.

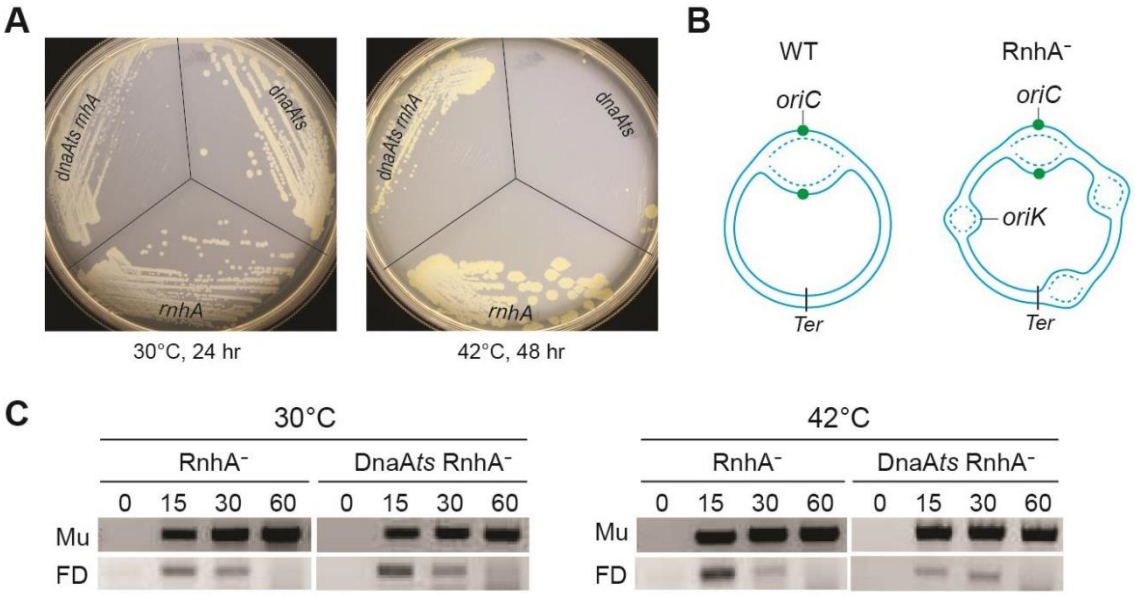


Figure 4.8 *oriC*-independent forks also promote FD repair.

(**A**) Growth of the *rnhA* (N4704), *dnaA46* (*ts*; AU1054), and *rnhA dnaA46* double mutant (AU1066) strains on LB agar plates at 30℃ and 42℃. The double mutant grows poorly even at 30℃. (**B**) Schematic depiction of replication originating at multiple ectopic locations in the RnhA mutant. (**C**) Mu integration and FD degradation in *rnhA* and *rnhA dnaA46* strains at 30℃ and 42℃. The latter infections were carried out after replication arrest for 90 min. The WT parent for these strains is MG1655, shown in Figure 4.2G.

Release of the replication block generates Mu-proximal DSBs

When the replication fork runs into the Mu transpososome, it is expected to stall not only because the Mu transpososome is extremely stable (Surette et al., 1987; Choi et al., 2014a), but also because the fork will eventually convert the gap adjacent to Mu into a DSB. To test for the presence of the expected DSB at the junction of Mu and the replication fork, we marked both Mu and the potential DSB with fluorescent markers. Mu was fluorescently labeled by first incorporating a TetO array into the prophage genome (Fig. 4.9A), such that phage viability was not affected; similar phage titers were obtained upon induction of both wild type and TetO-engineered prophages (Fig. 4.9B, C). When the Mu::TetO phage were infected into cells expressing TetR-mCherry at an MOI of 1, > 80% of the cells had single mCherry foci (Fig. 4.9D, E).

DSBs were fluorescently labeled using two different proteins, Mu Gam and RecA. Mu Gam binds directly to a double strand end (d'Adda di Fagagna et al., 2003), while RecA binds to the single-strand region of a processed DSB (Cox, 2001; Dillingham and Kowalczykowski, 2008). Fluorescent fusions to these proteins bind DSBs *in vivo* (Renzette et al., 2005; Shee et al., 2013). Since overexpression of Gam interferes with repair of DSBs and cell viability (see Fig. 4.10), its expression was controlled by the tightly regulated rhamnose-inducible promoter in pGam-GFP, repressing it with 0.2% glucose until needed; rhamnose was added only 30 min before Mu infection. RecA-GFP is expressed constitutively from its normal chromosomal location.

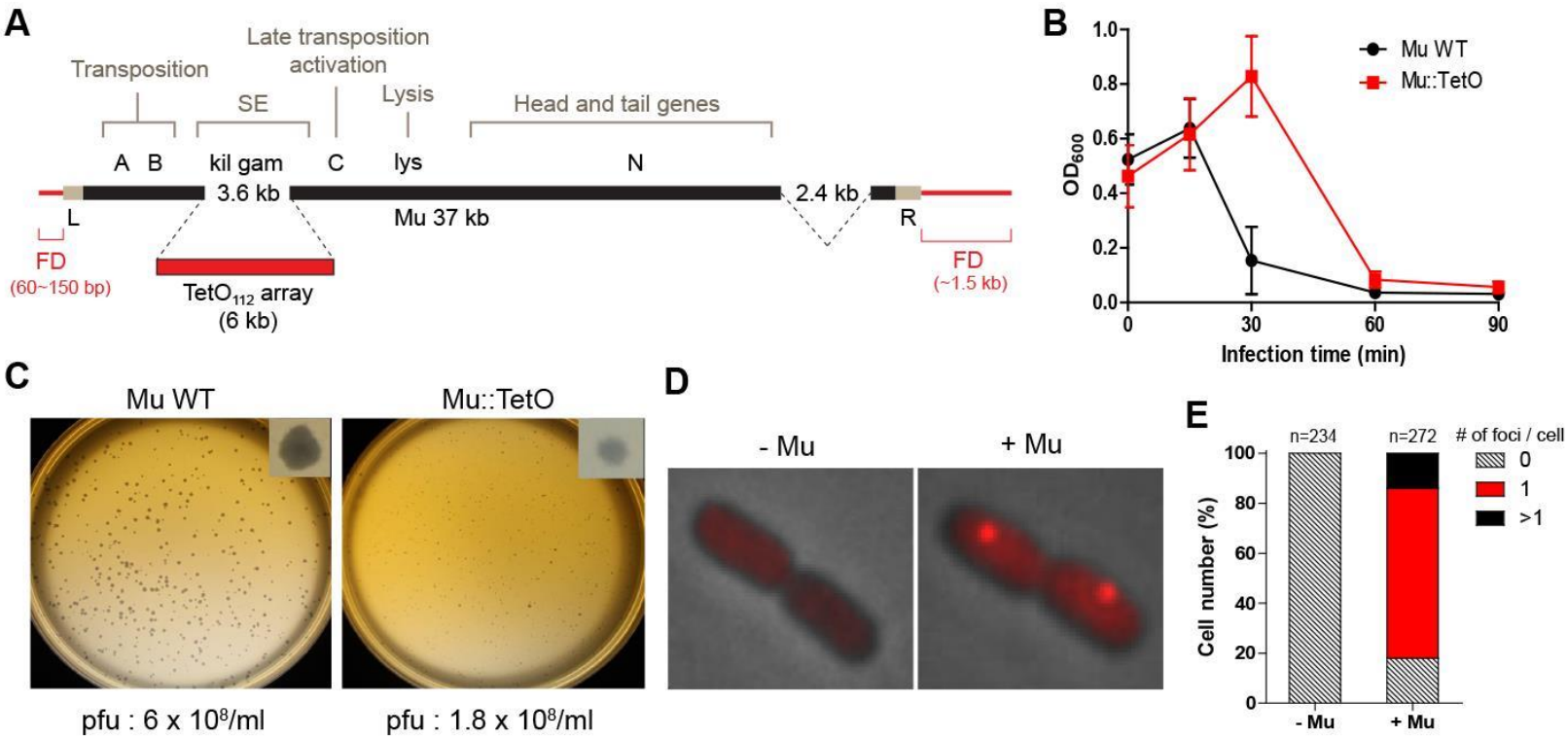


Figure 4.9 Visualizing integrated Mu with TetR-mCherry.

(**A**) The TetO array was substituted for the SE region of the prophage, which is dispensable for phage growth. A compensatory deletion in another non-essential region near the R end restored the original Mu DNA length, ensuring that head-full packaging would yield viable phage progeny (Symonds et al., 1987). See Materials and Methods for construction details. (**B**) Lysis

-Figure 4.9 *legend continued*-

profiles of Mu::TetO (SJ012) and Mu wild-type (HM8305) prophage strains. Similar phage titers were obtained from both strains (see C). Error bars indicate standard deviation from the mean of triplicate data sets obtained from three independent colonies of the same strain. (**C**) Plaque morphologies of wild-type and Mu::TetO, titered on BW25113. Mutations in the SE region are known to affect plaque size (inset) (Symonds et al., 1987). (**D**) Snapshot of Mu::TetO infection (MOI = 1) into WT strain BW25113 expressing TetR-mCherry (pDB317), photographed before (-Mu) and after ~15-20 min of infection (+Mu). (**E**) Quantitation of uninfected and infected cells in D by counting mCherry foci. > 80% of the cells had mCherry foci, and a majority of these had single foci.

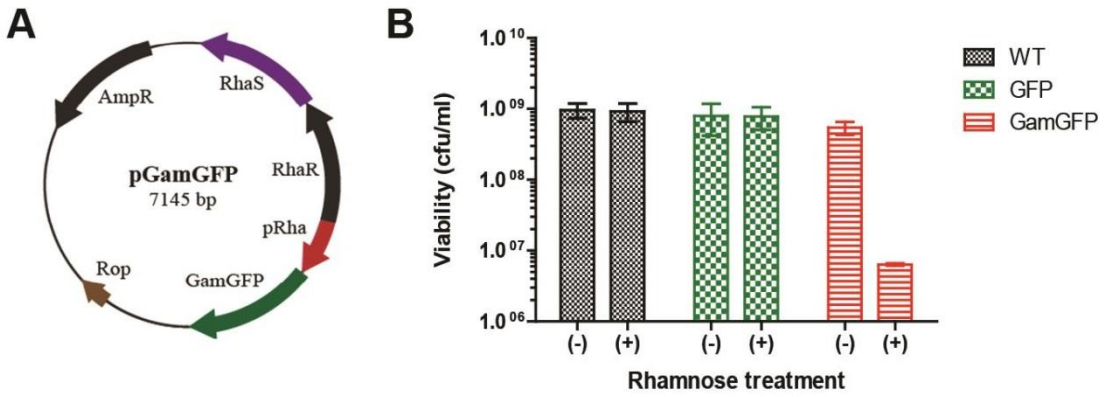


Figure 4.10 Gam-GFP expression and its effect on cell viability.

(**A**) Construction of the pGam-GFP expression vector. Details described under Materials and Methods. (**B**) Viability of WT cells (BW25113) expressing GFP (pGFP) or Gam-GFP (pGam-GFP). Overnight cultures of these strains propagated in LB with 0.2% glucose were diluted 1:100 into fresh LB medium and grown at 37℃ until the OD600 reached 0.2-0.3. They were grown for an additional 2 hours with or without 500 μM L-rhamnose prior to plating for cfu (colony forming units) on LB solid medium at 37℃. The data are from triplicate experimental repeats from three independent colonies of each strain. Error bars indicate standard deviation from the mean.

When cells expressing TetR-mCherry and Gam-GFP were infected with Mu::TetO at an MOI where the majority of cells had an mCherry focus (Fig.5), < 0.1% of the cells had visible GFP foci before infection; however, their numbers increased at least 30-fold after Mu infection (Fig. 4.11A). The green foci were binned into different categories with respect to their proximity to the red foci. Of the foci that appeared after Mu infection, ~ 40% either co-localized with or were proximal to Mu (Fig. 4.11A, right). When replication-arrested cells (DnaE*ts*) were infected with Mu, Gam-GFP foci appeared only when the replication block was removed (Fig. 4.11B and 4.12); 50% of these foci were associated with Mu. Foci not associated with Mu might be accounted for by the accumulated DNA damage in replication-arrested cells. The overall low numbers and low intensity of Gam-GFP foci may be due to limiting its expression to only 30 min prior to Mu infection, which was necessitated because prolonged expression had a negative impact on cell viability (Fig. 4.10).

When RecA-GFP was used to monitor DSBs, the GFP foci were bright and 24% of uninfected cells had a single focus, consistent with the reported spontaneous DSBs detected in cells growing in rich media (Fig. 4.11C) (Renzette et al., 2005). Infection with Mu increased their numbers by 13%. Here, >50% of the total number of the foci seen after infection were associated with Mu, suggesting that the majority of the Mu-induced RecA foci are near Mu. This experiment could not be conducted in a DnaE*ts* strain because of a high background of RecA-GFP foci at 42℃, likely due to induction of the SOS response in replication defective mutants (Renzette et al., 2005; Massoni and Sandler, 2013). That only a fraction of the Mu foci had associated RecA foci, could be due to fast repair of the DSB. This inference is supported by the observation that Mu infection does not induce the SOS response (Fig. 4.5D); induction of an SOS response requires that the DSB persist (Kuzminov, 1999; Lusetti and Cox, 2002). The independent

probability of localization of the GFP foci next to Mu-mCherry foci was estimated to be less than 0.6%, a number much lower than the observed 50% co-localization (see Fig. 4.11 legend). We conclude that Mu insertions generate DSBs in their vicinity that are dependent on the replication fork.

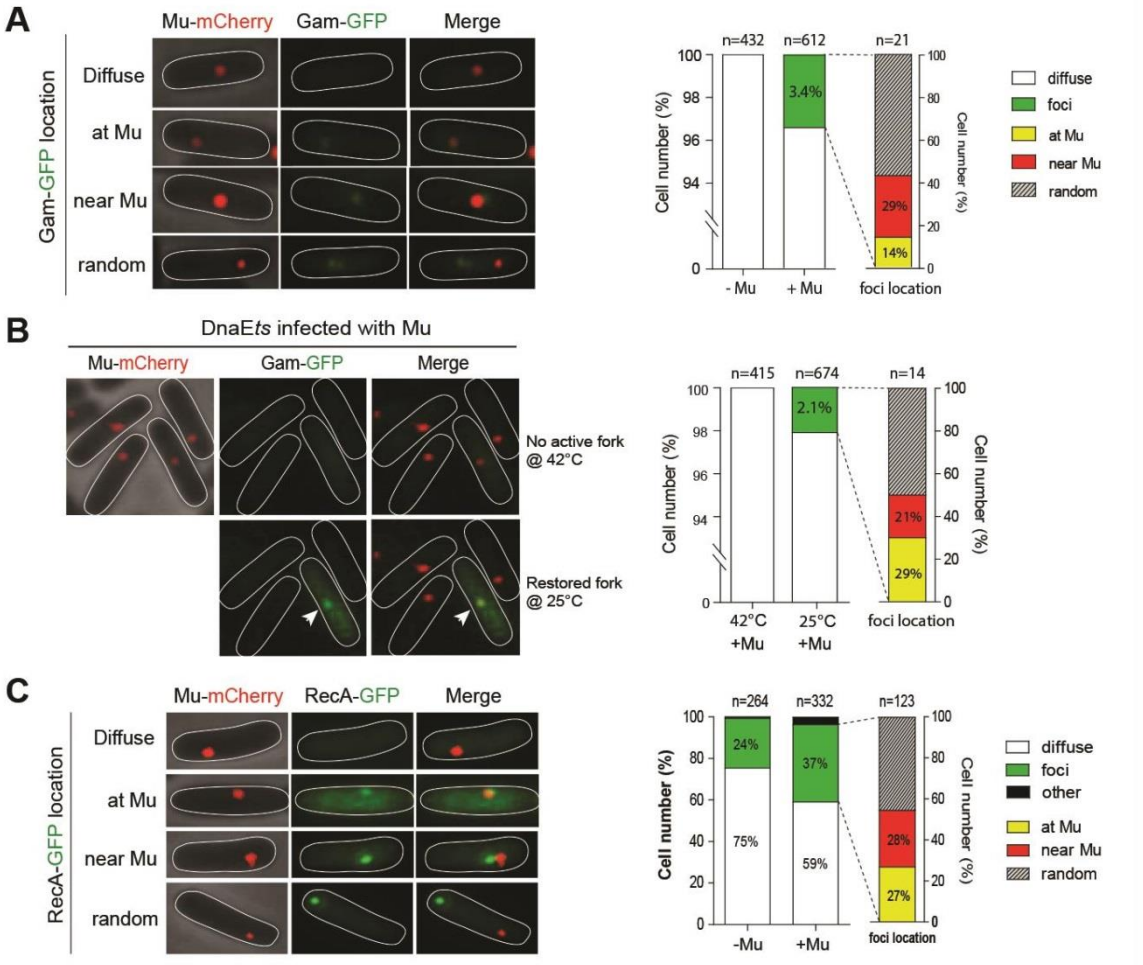


Figure 4.11 Mu insertions generate replication-dependent DSBs in their vicinity.

Mu::TetO is detected by TetR-mCherry (labeled Mu-mCherry) and DSBs by either Gam-GFP or RecA-GFP in WT (BW25113) or its DnaE*ts* derivative (SJ005). (**A**) Left, snapshot of Mu-mCherry foci relative to Gam-GFP foci upon Mu infection of WT. Right, quantitation of the position of the green and red foci. GFP foci were scored as 'at Mu' when they completely overlapped with mCherry foci, 'near Mu' when at least their edges touched, and 'random' when they did not touch. Background fluorescence with no detectable foci was scored 'diffuse'. (**B**) Left, replication-arrested cells were infected with Mu and monitored for Mu-mCherry and Gam-GFP foci (top panel). The same frame

-Figure 4.11 *legend continued*-

of cells was photographed after releasing the replication block by shifting cells to 25℃ (bottom panel). White arrow points to GFP focus that appeared after the temperature shift. Right, quantitation of foci after fork release, as in A. See also Figure 4.12. (**C**) As in A, but with a RecA-GFP expressing strain. Foci with large or aberrant morphology are classified 'other'. Using microbeTracker software (Sliusarenko et al., 2011) the average area of a cell and of a fluorescent focus were calculated to be 1.78 μm^2 and 0.0675 μm^2 respectively, yielding a total of 26 independent sites that any focus can occupy in the cell. Two foci that touch occupy 4 times the area of one focus. The probability that any two foci will be stochastically near each other is then modeled by $4/26^2$ or slightly less than 0.6%, well below the experimentally observed 50%, which include foci that completely overlap. See Materials and Methods for details.

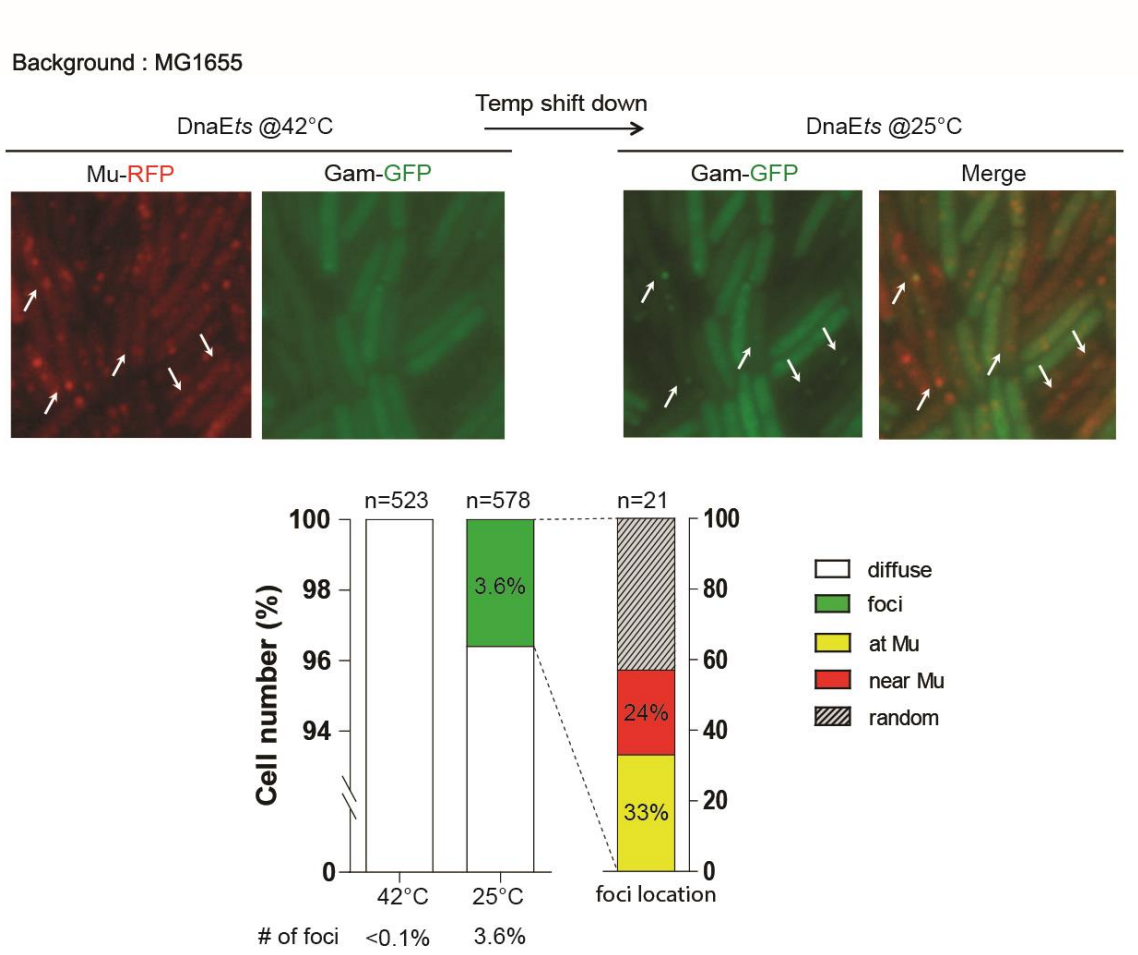


Figure 4.12 Replication-dependent DSBs proximal to Mu in a second host strain.

All descriptions as in Figure 4.11, except that the infected host strain was a derivative of MG1655 (SJ004). Top left panels: snapshot of Mu-mCherry and Gam-GFP foci in replication arrested cells. Top right panels: the same frame of cells was photographed after releasing the replication block. White arrows point to GFP foci that appeared after the temperature shift down. Bottom, quantitation of foci as described in Figure 4.11.

DISCUSSION

This work opens a new window into the repair of transposon insertions, the majority of which share with Mu the problem of repair of flanking target gaps. Progress in this area of transposon research has languished likely because of the difficulty of studying the repair of such short gaps. Mu has provided the first insight into this problem primarily because of the unusual feature of long FD sequences linked to the infecting Mu genome, which are repaired concomitant with gap repair (Fig. 4.2A). The development of a convenient *in vivo* assay for monitoring FD repair (Au et al., 2006), has allowed us to make large strides into understanding the repair process, both *in vivo* and *in vitro*. *In vivo*, the RecBCD exonuclease was found to be responsible for removal of the FD only after Mu integration. Thus, some post-integration event(s) must allow RecBCD access by releasing the N protein protecting the FD (Choi et al., 2014a). *In vitro*, a strand transfer substrate with FD lengths similar to the *in vivo* substrate but lacking N, showed that RecBCD degrades the FD until it encounters the transpososome, leaving +19 FD, which is converted to +4 FD with added cell extracts (Choi et al., 2014a). The present study confirms the generation of the short FD product *in vivo* (Figs. 4.3 and 4.4). The major finding of this study is that the event that triggers post-integration FD removal is arrival of the native replication fork at the Mu insertion site (Fig. 4.2-4.8). Thus, the transpososome teams up with the replisome to promote N removal. A requirement for the native fork explains why the DSB repair machinery is required to recover Mu insertions (Jang et al., 2012), because replication forks convert gaps into breaks, which must be repaired. We show that fork-dependent DSBs appear proximal to Mu (Fig. 4.11), and establish that gap-filling polymerases are not required for Mu repair (Fig. 4.5).

Model for Pol III-mediated coordinated repair of flanking target gaps without replicating the intervening Mu DNA

Why does Mu wait for the replisome and why does it employ a repair scheme that generates additional DNA damage for repair? We suggest that the Mu transpososome partners with the replisome for multiple tasks. First, the transpososome uses the incoming Pol III to signal just-in-time FD degradation, protecting the flanking gaps until the polymerase is on site to fill these gaps. Next, Pol III stalled at the DSB is exploited for coordinated repair of both target gaps flanking Mu without replicating the intervening 37 kb of Mu. Finally, the replisome is used for transpososome disassembly concomitant with gap repair, as discussed below.

According to the data in Figures 4.2-4.12, replisome interaction with the transpososome is somehow sensed at the distant N protein, destabilizing/removing N to allow RecBCD entry (Fig. 4.13A). While the FD is being shortened, Pol III is also in position for its leading strand subunit to fill the proximal Mu gap. However, the fork stalls, both because the transpososome is blocking its path and because a collateral DSB is formed on the lagging strand. The Pol III subunit on the broken lagging strand gets evicted while remaining attached to the leading strand subunit via the clamp loader (Fig. 4.13B). The evicted subunit is now available to reload on the target DNA flanking the distal Mu gap, which although 37 kb downstream, is actually proximal to Pol III because of a hairpin bend (140°) in the target DNA held within the transpososome (Montano et al., 2012). As the polymerase moves forward to fill both gaps, it must dislodge the transpososome because the gaps are protected within it (Lavoie et al., 1991; Mizuuchi et al., 1991). Such a replisome-transpososome co-operation ensures that the gaps are not exposed prior to transpososome disassembly, that gap repair is synchronized at both ends, and that gaps are filled without having to replicate the entire length of the intervening Mu

DNA (non-replicative transposition). The remaining +4 FD must be trimmed by a nuclease prior to sealing the DNA with ligase. The lagging strand DSB is subsequently repaired by homologous recombination, and the stalled replication fork reinstated by the Restart machinery (Fig. 4.13B) (Jang et al., 2012).

Why has Mu evolved an elaborate mechanism for removal and repair of the FD given that it can bypass FD repair to enter lytic growth in DnaA, RecBCD and some MuA mutants (this work and (Choi and Harshey, 2010; Choi et al., 2014a)). We believe that a mechanism limiting Mu replication to short gaps at the Mu ends gives the phage a chance to enter the prophage state.

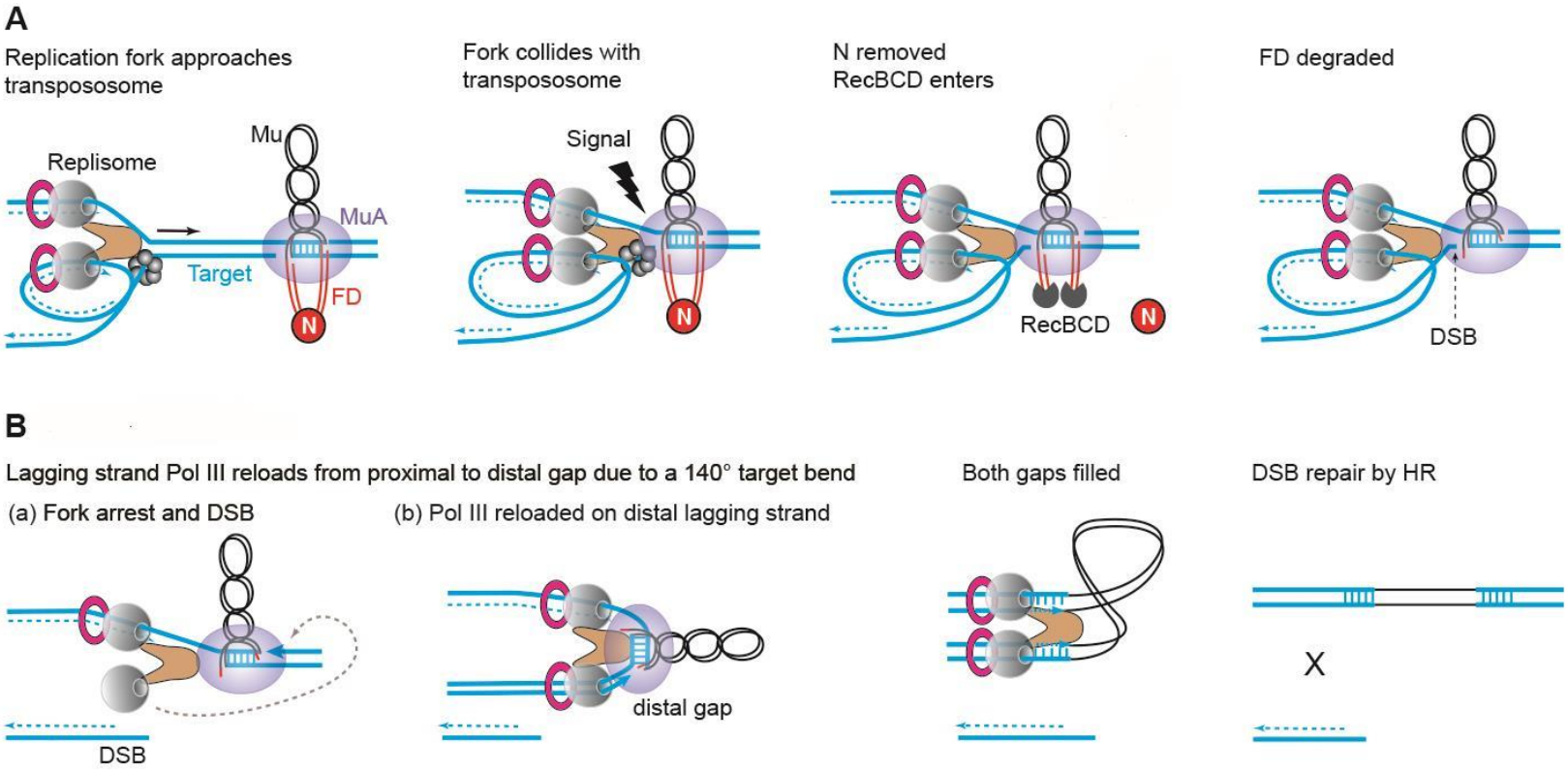


Figure 4.13 Model depicting how the Mu transpososome first exploits incoming Pol III for just-in-time FD degradation, and next exploits Pol III stalled at the DSB for coordinated repair of both target gaps flanking Mu.

(A) Sequence of repair events as deduced in this, and in prior work (Choi and Harshey, 2010; Choi et al., 2014a). The Mu transpososome does not begin FD repair until the replication fork arrives. Interaction between the two complexes generates a

- Figure 4.13 *legend continued*-

'signal' for N removal and RecBCD entry. The fork stalls because the transpososome is blocking its progress and also because a DSB is generated on the lagging strand. (**B**) Scheme for how the DSB and the bent target is exploited for repair of both gaps during non-replicative Mu transposition. *E. coli* replisomes contain three polymerase molecules (Reyes-Lamothe et al., 2010), but only two are shown for clarity. The Pol III subunit evicted by the DSB (or the extra third subunit) reloads on the distal gap brought into proximity by the hairpin target bend within the transpososome (Montano et al., 2012). As the polymerase moves forward, the transpososome is dislodged and both gaps are filled simultaneously. A nuclease must trim the +4 FD, and a ligase seal the remaining nicks. The DSB is repaired by homologous recombination (HR).

Repair of other transposon insertions

The short gaps generated in the target are a universal feature of transposition. The +4 Mu overhangs are similar to +3/+2 overhangs on the 5' flanking DNA of transposition intermediates of Tn7 and retroviral-like transposons (Bainton et al., 1991; Craigie, 2002; Sandmeyer et al., 2002). Although generated by different mechanisms in each element, the overhangs will encounter the common fate of being removed during repair of the gaps. Besides a common structure shared by the strand transfer intermediates of Mu and retroviral elements, the transpososomes of these elements also share many structural features, including their extraordinary stability and a bent target conformation (Maertens et al., 2010; Montano et al., 2012). The bent target positions two closely spaced target phosphodiester bonds for nucleophilic attack from the 3'OHs of the cleaved transposon ends. This chemistry of transposition is shared by all transposons, so it is likely that the target bend is also a shared by all transpososomes (Chandler and Craig, 2015; Harshey, 2015). While Mu has an elaborate mechanism to disassemble the transpososome during transition to replication (Fig. 4.1), the results in this paper suggest that the replisome might participate in transpososome disassembly during non-replicative transposition, which is the most prevalent transposition mechanism. *In vitro* attempts to recapitulate the gap-filling reaction after retroviral transposition has identified several polymerases, ligase and FEN-1 nuclease, but the *in vivo* chromatin substrate is substantially different, and the participating enzymes are likely to be different (Daniel, 2006). Interestingly, like with Mu, DSB repair proteins have been implicated in repair of retroviral integration events (Daniel, 2006; Yang et al., 2010). All these shared features hint at a common pathway of gap repair in all transposons.

PERSPECTIVE

Mu has played a central role in the development of the mobile DNA element field (Harshey, 2012). Pioneering *in vitro* Mu experiments led to the unraveling of the phosphoryl transfer chemistry for all transposable elements (Mizuuchi and Baker, 2002). High-throughput integration assays modeled after Mu (Craigie et al., 1991), led to the development and marketing of the HIV integrase inhibitor Raltegravir (Summa et al., 2008). We expect our *in vivo* study on the repair of Mu insertions to provide equally important insights into the long-standing problem of how DNA transposons, retrotransposons, and retroviruses disassemble their highly stable transpososome intermediates and repair their insertions without duplicating themselves in the process, presenting perhaps a new target for drug development.

Beyond repair of transposition events, these results may be relevant to replication-dependent repair of DNA lesions such as the interstrand crosslinks that produce DNA kinks/bends, where gap repair is associated with lesion bypass, an area under active investigation (Ho and Scharer, 2010; McVey, 2010; Pathania et al., 2011).

References

Abdelhakim, A. H., et al. (2008). "Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA plus unfoldase ClpX." *Mol Cell* **30**(1): 39-50.

Abdelhakim, A. H., et al. (2010). "The AAA+ ClpX machine unfolds a keystone subunit to remodel the Mu transpososome." *Proc Natl Acad Sci U S A* **107**(6): 2437-2442.

Akroyd, J. E. and N. Symonds (1983). "Evidence for a conservative pathway of transposition of bacteriophage Mu." *Nature* **303**(5912): 84-86.

Aldaz, H., et al. (1996). "The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis." *Cell* **85**(2): 257-269.

Au, T. K., et al. (2006). "Chromosomal integration mechanism of infecting mu virion DNA." *J Bacteriol* **188**(5): 1829-1834.

Ausubel, F. M. and *e. al* (2003). Current Protocols in Molecular Biology. New York, NY. , Jon Wiley & Sons Inc. .

Baba, T., et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." *Mol Syst Biol* **2**: 2006 0008.

Bainton, R., et al. (1991). "Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA." *Cell* **65**: 805-816.

Barnes, D. E. and T. Lindahl (2004). "Repair and genetic consequences of endogenous DNA base damage in mammalian cells." *Annu Rev Genet* **38**: 445-476.

Blattner, F. R., et al. (1997). "The complete genome sequence of Escherichia coli K-12." *Science* **277**(5331): 1453-1462.

Boonsombat, R., et al. (2006). "A novel dnaC mutation that suppresses priB rep mutant phenotypes in Escherichia coli K-12." *Mol Microbiol* **60**(4): 973-983.

Bork, J. M., et al. (2001). "The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA." *Embo J* **20**(24): 7313-7322.

Bujacz, G., et al. (1996). "The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations." *Structure* **4**(1): 89-96.

Bukhari, A. I. (1975). "Reversal of mutator phage Mu integration." *J Mol Biol* **96**: 87-99.

Bukhari, A. I. and A. L. Taylor (1975). "Influence of insertions on packaging of host sequences covalently linked to bacteriophage Mu DNA." *Proc Natl Acad Sci U S A* **72**(11): 4399-4403.

Chaconas, G., et al. (1985). "A truncated form of the bacteriophage Mu B protein promotes conservative integration, but not replicative transposition, of Mu DNA." *Cell* **41**(3): 857-865.

Chaconas, G., et al. (1984). "Transposition of bacteriophage mu DNA: expression of the A and B proteins from lambda pL and analysis of infecting mu DNA." *Cold Spring Harb Symp Quant Biol* **49**: 279-284.

Chaconas, G. and R. M. Harshey (2002). Transposition of phage Mu DNA. Washington DC, ASM Press.

Chaconas, G., et al. (1981). "Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions." *J Mol Biol* **150**(3): 341-359.

Chaconas, G., et al. (1983). "Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand." *Virology* **128**(1): 48-59.

Chandler, M. and N. Craig, L. (2015). Mobile DNA III. Washington D. C., ASM.

Chen, J. C., et al. (2000). "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding." *Proc Natl Acad Sci U S A* **97**(15): 8233-8238.

Chenais, B. (2013). "Transposable elements and human cancer: a causal relationship?" *Biochim Biophys Acta* **1835**(1): 28-35.

Choi, W. and R. M. Harshey (2010). "DNA repair by the cryptic endonuclease activity of Mu transposase." *Proc Natl Acad Sci U S A* **107**(22): 10014-10019.

Choi, W., et al. (2014a). "Mu transpososome and RecBCD nuclease collaborate in the repair of simple Mu insertions." *Proc Natl Acad Sci U S A* **111**(39): 14112-14117.

Choi, W., et al. (2014b). "Controlling DNA degradation from a distance: a new role for the Mu transposition enhancer." *Mol Microbiol* **94**(3): 595-608.

Clements, M. B. and M. Syvanen (1981). "Isolation of polA mutation that affects transposition of insertion sequences and transposons." *Cold Spring Harb Symp Quant Biol* **45 Pt 1**: 201-204.

Coffin, J. M. (1992). "Retroviral DNA integration." *Dev Biol Stand* **76**: 141-151.

Cooper, S. and C. E. Helmstetter (1968). "Chromosome replication and the division cycle of *Escherichia coli* B/r." *J Mol Biol* **31**: 519-540.

Cox, M. M. (2001). "Recombinational DNA repair of damaged replication forks in Escherichia coli: questions." *Annu Rev Genet* **35**: 53-82.

Craig, N., L., Craigie, R., Gellert, M and Lambowitz, A (2002). Mobile DNA II. Washington, D.C., ASM Press.

Craig, N. L. (1996). "Transposon Tn7." *Curr Top Microbiol Immunol* **204**: 27-48.

Craigie, R. (2002). Retroviral DNA integration. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington, DC, ASM Press**:** 613-630.

Craigie, R., et al. (1991). "A rapid in vitro assay for HIV DNA integration." *Nucleic Acids Res* **19**(10): 2729-2734.

Curcio, M. J. and K. M. Derbyshire (2003). "The outs and ins of transposition: from mu to kangaroo." *Nat Rev Mol Cell Biol* **4**(11): 865-877.

d'Adda di Fagagna, F., et al. (2003). "The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku." *EMBO Rep* **4**: 47-52.

Daniel, R. (2006). "DNA repair in HIV-1 infection: a case for inhibitors of cellular co-factors?" *Curr HIV Res* **4**(4): 411-421.

Daniel, R., et al. (2003). "Evidence that the retroviral DNA integration process triggers an ATR-dependent DNA damage response." *Proc Natl Acad Sci U S A* **100**(8): 4778-4783.

Darmon, E. and D. R. Leach (2014). "Bacterial genome instability." *Microbiol Mol Biol Rev* **78**(1): 1-39.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." *Proc Natl Acad Sci U S A* **97**(12): 6640-6645.

Dewannieux, M., et al. (2003). "LINE-mediated retrotransposition of marked Alu sequences." *Nat Genet* **35**(1): 41-48.

Dillingham, M. S. and S. C. Kowalczykowski (2008). "RecBCD enzyme and the repair of double-stranded DNA breaks." *Microbiol Mol Biol Rev* **72**(4): 642-671, Table of Contents.

Drake, J. W., et al. (1998). "Rates of spontaneous mutation." *Genetics* **148**(4): 1667-1686.

Engelman, A., et al. (1991). "HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer." *Cell* **67**(6): 1211-1221.

Espeseth, A. S., et al. (2011). "siRNA screening of a targeted library of DNA repair factors in HIV infection reveals a role for base excision repair in HIV integration." *plos one* **6**(3): e17612.

Fedoroff, N. V. (1994). "Barbara McClintock (June 16, 1902-September 2, 1992)." *Genetics* **136**(1): 1-10.

Fedoroff, N. V. (2012). "Presidential address. Transposable elements, epigenetics, and genome evolution." *Science* **338**(6108): 758-767.

Fijalkowska, I. J., et al. (2012). "DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair." *FEMS Microbiol Rev* **36**: 1105-1121.

Gabbai, C. B. and K. J. Marians (2010). "Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival." *DNA Repair (Amst)* **9**(3): 202-209.

Gasior, S. L., et al. (2006). "The human LINE-1 retrotransposon creates DNA double-strand breaks." *J Mol Biol* **357**(5): 1383-1393.

Ge, J., et al. (2011). "Analysis of phage Mu DNA transposition by whole-genome Escherichia coli tiling arrays reveals a complex relationship to distribution of target selection protein B, transcription and chromosome architectural elements." *J Biosci* **36**(4): 587-601.

George, M. and A. I. Bukhari (1981). "Heterogeneous host DNA attached to the left end of mature bacteriophage Mu DNA." *Nature* **292**: 175-176.

Giacalone, M. J., et al. (2006). "Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system." *Biotechniques* **40**(3): 355-364.

Gloor, G. and G. Chaconas (1986). "The bacteriophage Mu N gene encodes the 64-kDa virion protein which is injected with, and circularizes, infecting Mu DNA." *J Biol Chem* **261**(35): 16682-16688.

Goodman, M. F. (2002). "Error-prone repair DNA polymerases in prokaryotes and eukaryotes." *Annu Rev Biochem* **71**: 17-50.

Grindley, N. D. F. (1983). "Transposition of Tn3 and Related Transposons." *Cell* **32**(1): 3-5.

Gueguen, E., et al. (2005). "The transpososome: control of transposition at the level of catalysis." *Trends Microbiol* **13**(11): 543-549.

Gumbiner-Russo, L. M. and S. M. Rosenberg (2007). "Physical analyses of E. coli heteroduplex recombination products in vivo: on the prevalence of 5' and 3' patches." *PLoS One* **2**(11): e1242.

Hare, S., et al. (2010). "Retroviral intasome assembly and inhibition of DNA strand transfer." *Nature* **464**(7286): 232-236.

Haren, L., et al. (1999). "Integrating DNA: transposases and retroviral integrases." *Annu Rev Microbiol* **53**: 245-281.

Harshey, R. M. (1984). "Transposition without Duplication of Infecting Bacteriophage Mu DNA." *Nature* **311**(5986): 580-581.

Harshey, R. M. (2012). "The Mu story: how a maverick phage moved the field forward." *Mobile DNA* **3**: 21.

Harshey, R. M. (2015). Transposable phage Mu. Mobile DNA III. M. Chandler and N. L. Craig. Washington, D.C., ASM press**:** In press.

Harshey, R. M. and A. I. Bukhari (1983). "Infecting bacteriophage mu DNA forms a circular DNA-protein complex." *J Mol Biol* **167**(2): 427-441.

Higgins, N. P., et al. (1989). "Supercoiling and integration host factor change the DNA conformation and alter the flow of convergent transcription in phage Mu." *J Biol Chem* **264**(5): 3035-3042.

Hiom, K., et al. (1998). "DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations." *Cell* **94**(4): 463-470.

Ho, T. V. and O. D. Scharer (2010). "Translesion DNA synthesis polymerases in DNA interstrand crosslink repair." *Environ Mol Mutagen* **51**(6): 552-566.

Horvath, P. and R. Barrangou (2010). "CRISPR/Cas, the immune system of bacteria and archaea." *Science* **327**(5962): 167-170.

Howe, M. M. and E. G. Bade (1975). "Molecular biology of bacteriophage mu." *Science* **190**(4215): 624-632.

Huang, C. R., et al. (2012). "Active transposition in genomes." *Annu Rev Genet* **46**: 651-675.

Hughes, K. T., et al. (1987). "Rec dependence of mu transposition from P22-transduced fragments." *J Bacteriol* **169**(1): 403-409.

Jackson, S. P. and J. Bartek (2009). "The DNA-damage response in human biology and disease." *Nature* **461**(7267): 1071-1078.

Jang, S., et al. (2012). "Mu insertions are repaired by the double-strand break repair pathway of *Escherichia coli.*" *PLoS genetics* **8**: e1002642.

Jones, J. M. and H. Nakai (1997). "The phiX174-type primosome promotes replisome assembly at the site of recombination in bacteriophage Mu transposition." *EMBO J* **16**(22): 6886-6895.

Jones, J. M. and H. Nakai (1999). "Duplex opening by primosome protein PriA for replisome assembly on a recombination intermediate." *J Mol Biol* **289**(3): 503-516.

Jones, J. M. and H. Nakai (2000). "PriA and phage T4 gp59: factors that promote DNA replication on forked DNA substrates MicroReview [In Process Citation]." *Mol Microbiol* **36**(3): 519-527.

Joshi, M. C., et al. (2013). "Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA." *PLoS Genet* **9**(8): e1003673.

Joyce, C. M. and N. D. Grindley (1984). "Method for determining whether a gene of Escherichia coli is essential: application to the polA gene." *J Bacteriol* **158**(2): 636-643.

Kasuga, T. and M. Gijzen (2013). "Epigenetics and the evolution of virulence." *Trends Microbiol* **21**(11): 575-582.

Kazazian, H. H., Jr. (1999). "An estimated frequency of endogenous insertional mutations in humans." *Nat Genet* **22**(2): 130.

Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." *Science* **303**(5664): 1626-1632.

Kennedy, A. K., et al. (1998). "Tn10 transposition via a DNA hairpin intermediate." *Cell* **95**(1): 125-134.

Kidwell, M. G. and D. R. Lisch (2000). "Transposable elements and host genome evolution." *Trends Ecol Evol* **15**(3): 95-99.

Kim, M. S., et al. (2015). "Crystal structure of the V(D)J recombinase RAG1-RAG2." *Nature* **518**(7540): 507-511.

Kogoma, T. (1997). "Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription." *Microbiol Mol Biol Rev* **61**(2): 212-238.

Konrad, E. B. (1977). "Method for the isolation of *Escherichia coli* mutants with enhanced recombination between chromosomal duplications." *J Bacteriol* **130**: 167-172.

Kornberg, A. and T. A. Baker (1992). DNA Replication, W. H. Freeman & Co.

Kruklitis, R. and H. Nakai (1994). "Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis in vitro." *J Biol Chem* **269**(23): 16469-16477.

Kruklitis, R., et al. (1996). "ClpX protein of Escherichia coli activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis." *EMBO J* **15**(4): 935-944.

Kunarso, G., et al. (2010). "Transposable elements have rewired the core regulatory network of human embryonic stem cells." *Nat Genet* **42**(7): 631-634.

Kuzminov, A. (1999). "Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda." *Microbiol Mol Biol Rev* **63**(4): 751-813, table of contents.

Lander, E. S., et al. (2001). "Initial sequencing and analysis of the human genome." *Nature* **409**(6822): 860-921.

Lauder, S. D. and S. C. Kowalczykowski (1993). "Negative co-dominant inhibition of recA protein function. Biochemical properties of the recA1, recA13 and recA56 proteins and the effect of recA56 protein on the activities of the wild-type recA protein function in vitro." *J Mol Biol* **234**(1): 72-86.

Lavoie, B. D. and G. Chaconas (1996). "Transposition of phage Mu DNA." *Transposable Elements* **204**: 83-102.

Lavoie, B. D., et al. (1991). "Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome." *EMBO J* **10**(10): 3051-3059.

Lee, E. H. and A. Kornberg (1991). "Replication deficiencies in priA mutants of Escherichia coli lacking the primosomal replication n' protein." *Proc Natl Acad Sci U S A* **88**(8): 3029-3032.

Levchenko, I., et al. (1995). "Disassembly of the Mu transposase tetramer by the ClpX chaperone." *Genes Dev* **9**(19): 2399-2408.

Li, L., et al. (2001). "Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection." *EMBO J* **20**(12): 3272-3281.

Li, W., et al. (2013). "Activation of transposable elements during aging and neuronal decline in Drosophila." *Nat Neurosci* **16**(5): 529-531.

Liebart, J. C., et al. (1982). "Conservative integration of bacteriophage Mu DNA into pBR322 plasmid." *Proc Natl Acad Sci U S A* **79**(14): 4362-4366.

Little, J. W., et al. (1980). "Cleavage of the Escherichia coli lexA protein by the recA protease." *Proc Natl Acad Sci U S A* **77**(6): 3225-3229.

Liu, J., et al. (1999). "Replication fork assembly at recombination intermediates is required for bacterial growth." *Proc Natl Acad Sci U S A* **96**(7): 3552-3555.

Lusetti, S. L. and M. M. Cox (2002). "The bacterial RecA protein and the recombinational DNA repair of stalled replication forks." *Annu Rev Biochem* **71**: 71-100.

Maduike, N. Z., et al. (2014). "Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics." *Mol Microbiol* **91**(1): 39-56.

Maertens, G. N., et al. (2010). "The mechanism of retroviral integration from X-ray structures of its key intermediates." *Nature* **468**(7321): 326-329.

Mahillon, J. and M. Chandler (1998). "Insertion sequences." *Microbiol Mol Biol Rev* **62**(3): 725-774.

Manna, D., et al. (2004). "Microarray analysis of transposition targets in Escherichia coli: the impact of transcription." *Proc Natl Acad Sci U S A* **101**(26): 9780-9785.

Massoni, S. C. and S. J. Sandler (2013). "Specificity in suppression of SOS expression by *recA4162* and *uvrD303*." *DNA Repair (Amst)* **12**: 1072-1080.

McBeth, D. L. and A. L. Taylor (1982). "Growth of bacteriophage Mu in Escherichia coli dnaA mutants." *J Virol* **44**(2): 555-564.

McBeth, D. L. and A. L. Taylor (1983). "Involvement of Escherichia coli K-12 DNA polymerase I in the growth of bacteriophage Mu." *J Virol* **48**(1): 149-156.

McClintock, B. (1950). "The origin and behavior of mutable loci in maize." *Proc Natl Acad Sci U S A* **36**(6): 344-355.

McCool, J. D., et al. (2004a). "A dnaT mutant with phenotypes similar to those of a priA2::kan mutant in Escherichia coli K-12." *Genetics* **167**(2): 569-578.

McCool, J. D., et al. (2004b). "Measurement of SOS expression in individual Escherichia coli K-12 cells using fluorescence microscopy." *Mol Microbiol* **53**(5): 1343-1357.

McCool, J. D. and S. J. Sandler (2001). "Effects of mutations involving cell division, recombination, and chromosome dimer resolution on a priA2::kan mutant." *Proc Natl Acad Sci U S A* **98**(15): 8203-8210.

McGlynn, P. and R. G. Lloyd (2002). "Recombinational repair and restart of damaged replication forks." *Nat Rev Mol Cell Biol* **3**(11): 859-870.

McVey, M. (2010). "Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds." *Environ Mol Mutagen* **51**(6): 646-658.

Mhammedi-Alaoui, A., et al. (1994). "A new component of bacteriophage Mu replicative transposition machinery: the Escherichia coli ClpX protein." *Mol Microbiol* **11**(6): 1109-1116.

Michaelis, C., et al. (1997). "Cohesins: chromosomal proteins that prevent premature separation of sister chromatids." *Cell* **91**(1): 35-45.

Michel, B., et al. (2007). "Recombination proteins and rescue of arrested replication forks." *DNA Repair (Amst)* **6**(7): 967-980.

Michel, B., et al. (2004). "Multiple pathways process stalled replication forks." *Proc Natl Acad Sci U S A* **101**(35): 12783-12788.

Miller, J. H. (1992). A short course in bacterial genetics. Cold Spring Harbor, NY., Cold Spring Harbor Laboratory Press

Mills, R. E., et al. (2007). "Which transposable elements are active in the human genome?" *Trends Genet* **23**(4): 183-191.

Mizuuchi, K. (1983). "In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction." *Cell* **35**(3 Pt 2): 785-794.

Mizuuchi, K. (1984). "Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition." *Cell* **39**(2 Pt 1): 395-404.

Mizuuchi, K. (1992). "Polynucleotidyl transfer reactions in transpositional DNA recombination." *J Biol Chem* **267**(30): 21273-21276.

Mizuuchi, K. (1992). "Transpositional recombination: mechanistic insights from studies of Mu and other elements." *Annu Rev Biochem* **61**: 1011-1051.

Mizuuchi, K. and T. A. Baker (2002). Chemical mechanisms for mobilizing DNA. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM Press**:** 12-23.

Mizuuchi, M., et al. (1991). "DNase protection analysis of the stable synaptic complexes involved in Mu transposition." *Proc Natl Acad Sci U S A* **88**(20): 9031-9035.

Mizuuchi, M. and K. Mizuuchi (1993). "Target site selection in transposition of phage Mu." *Cold Spring Harb Symp Quant Biol* **58**: 515-523.

Mizuuchi, M. and K. Mizuuchi (2001). "Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB." *EMBO J* **20**(23): 6927-6935.

Montano, S. P., et al. (2012). "The Mu transpososome structure sheds light on DDE recombinase evolution." *Nature* **491**: 413-417.

Morgan, G. J., et al. (2002). "Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus." *J Mol Biol* **317**(3): 337-359.

Muller, K. H., et al. (1988). "Unmasking of bacteriophage Mu lipopolysaccharide receptors in Salmonella enteritidis confers sensitivity to Mu and permits Mu mutagenesis." *J Bacteriol* **170**(3): 1076-1081.

Nakai, H., et al. (2001). "Handoff from recombinase to replisome: insights from transposition." *Proc Natl Acad Sci U S A* **98**(15): 8247-8254.

Nakai, H. and A. L. Taylor (1985). "Host DNA replication forks are not preferred targets for bacteriophage Mu transposition." *J Bacteriol* **163**(1): 282-290.

Namgoong, S. Y. and R. M. Harshey (1998). "The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition." *EMBO J* **17**(13): 3775-3785.

North, S. H., et al. (2007). "Translation factor IF2 at the interface of transposition and replication by the PriA-PriC pathway." *Mol Microbiol* **66**(6): 1566-1578.

North, S. H. and H. Nakai (2005). "Host factors that promote transpososome disassembly and the PriA-PriC pathway for restart primosome assembly." *Mol Microbiol* **56**(6): 1601-1616.

Nurse, P., et al. (1991). "Inactivation of the Escherichia coli priA DNA replication protein induces the SOS response." *J Bacteriol* **173**(21): 6686-6693.

O'Day, K. J., et al. (1978). A search for integration deficient mutants of bacteriophage Mu-1. Microbiology. D. Schlessinger. Washington, D.C., ASM Publications**:** 48-51.

Ohtsubo, E., et al. (1981). "Mechanism of insertion and cointegration mediated by IS1 and Tn3." *Cold Spring Harb Symp Quant Biol* **45 Pt 1**: 283-295.

Papamichos-Chronakis, M. and C. L. Peterson (2013). "Chromatin and the genome integrity network." *Nat Rev Genet* **14**(1): 62-75.

Parsons, R. L. and R. Harshey (1988). "Autoregulation of phage Mu transposase at the level of translation." *Nuc Acids Res* **16**: 11285-11301.

Pathania, S., et al. (2011). "BRCA1 is required for postreplication repair after UV-induced DNA damage." *Mol Cell* **44**(2): 235-251.

Pato, M. L. (2004). "Replication of Mu prophages lacking the central strong gyrase site." *Res Microbiol* **155**: 553-558.

Perrat, P. N., et al. (2013). "Transposition-driven genomic heterogeneity in the Drosophila brain." *Science* **340**(6128): 91-95.

Poirier, M. C. (2004). "Chemical-induced DNA damage and human cancer risk." *Nat Rev Cancer* **4**(8): 630-637.

Ponder, R. G., et al. (2005). "A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation." *Mol Cell* **19**(6): 791-804.

Prak, E. T. and H. H. Kazazian, Jr. (2000). "Mobile elements and the human genome." *Nat Rev Genet* **1**(2): 134-144.

Pribil, P. A. and D. B. Haniford (2003). "Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition." *J Mol Biol* **330**(2): 247-259.

Puspurs, A. H., et al. (1983). "Bacteriophage Mu DNA circularizes following infection of Escherichia coli." *Embo J* **2**(3): 345-352.

Register, J. C., 3rd and J. Griffith (1985). "The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange." *J Biol Chem* **260**(22): 12308-12312.

Renzette, N., et al. (2005). "Localization of RecA in Escherichia coli K-12 using RecA-GFP." *Mol Microbiol* **57**(4): 1074-1085.

Reyes-Lamothe, R., et al. (2010). "Stoichiometry and architecture of active DNA replication machinery in Escherichia coli." *Science* **328**(5977): 498-501.

Rice, P. A. and T. A. Baker (2001). "Comparative architecture of transposase and integrase complexes." *Nat Struct Biol* **8**(5): 302-307.

Rudolph, C. J., et al. (2013). "Avoiding chromosome pathology when replication forks collide." *Nature* **500**(7464): 608-611.

Saluja, D. and G. N. Godson (1995). "Biochemical characterization of *Escherichia coli* temperature-sensitive *dnaB* mutants *dnaB8*, *dnaB252*, *dnaB70*, *dnaB43,* and *dnaB454*." *J Bacteriol* **177**: 1104-1111.

Sandler, S. J. (2005). "Requirements for replication restart proteins during constitutive stable DNA replication in Escherichia coli K-12." *Genetics* **169**(4): 1799-1806.

Sandler, S. J., et al. (1999). "dnaC mutations suppress defects in DNA replication- and recombination-associated functions in priB and priC double mutants in Escherichia coli K-12." *Mol Microbiol* **34**(1): 91-101.

Sandler, S. J., et al. (2001). "PriA mutations that affect PriA-PriC function during replication restart." *Mol Microbiol* **41**(3): 697-704.

Sandler, S. J., et al. (1996). "Differential suppression of priA2::kan phenotypes in Escherichia coli K-12 by mutations in priA, lexA, and dnaC." *Genetics* **143**(1): 5-13.

Sandmeyer, S. B., et al. (2002). Ty3, a position-specific, gypsy-like element in *Saccharomyces cerevisiae*. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington, DC, ASM Press**:** 663-683.

Sandulache, R., et al. (1984). "Cell wall receptor for bacteriophage Mu G(+)." *J Bacteriol* **160**(1): 299-303.

Santarius, T., et al. (2010). "A census of amplified and overexpressed human cancer genes." *Nat Rev Cancer* **10**(1): 59-64.

Sasakawa, C., et al. (1981). "The requirement for both DNA polymerase and 5' to 3' exonuclease activities of DNA polymerase I during Tn5 transposition." *Mol Gen Genet* **182**(1): 19-24.

Savilahti, H. and K. Mizuuchi (1996). "Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase." *Cell* **85**(2): 271-280.

Schnable, P. S., et al. (2009). "The B73 maize genome: complexity, diversity, and dynamics." *Science* **326**(5956): 1112-1115.

Shee, C., et al. (2013). "Engineered proteins detect spontaneous DNA breakage in human and bacterial cells." *Elife* **2**: e01222.

Singer, T., et al. (2010). "LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?" *Trends Neurosci* **33**(8): 345-354.

Sinha, R. P. and D. P. Hader (2002). "UV-induced DNA damage and repair: a review." *Photochem Photobiol Sci* **1**(4): 225-236.

Skalka, A. M. and R. A. Katz (2005). "Retroviral DNA integration and the DNA damage response." *Cell Death Differ* **12**: 971-978.

Skinner, S. O., et al. (2013). "Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization." *Nature protocols* **8**(6): 1100-1113.

Sliusarenko, O., et al. (2011). "High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics." *Mol Microbiol* **80**(3): 612-627.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." *Nat Rev Genet* **8**(4): 272-285.

Smith, J. A. and R. Daniel (2006). "Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses." *ACS Chem Biol* **1**(4): 217-226.

Sonti, R. V., et al. (1993). "Lethal transposition of Mud phages in Rec- strains of Salmonella typhimurium." *Genetics* **133**(1): 17-28.

Summa, V., et al. (2008). "Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection." *J Med Chem* **51**(18): 5843-5855.

Surette, M. G., et al. (1987). "Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA." *Cell* **49**: 253-262.

Suzuki, J., et al. (2009). "Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition." *PLoS Genet* **5**(4): e1000461.

Symonds, N., et al. (1987). Phage Mu. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.

Syvanen, M., et al. (1982). "A new class of mutants in DNA polymerase I that affects gene transposition." *J Mol Biol* **158**: 203-212.

Takeda, H., et al. (2015). "Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression." *Nat Genet* **47**(2): 142-150.

Taylor, A. L. (1963). "Bacteriophage-Induced Mutation in Escherichia Coli." *Proc Natl Acad Sci U S A* **50**: 1043-1051.

Toussaint, A. and M. Faelen (1974). "The dependence of temperate phage Mu-1 upon replication functions of E. coli K12." *Mol Gen Genet* **131**(3): 209-214.

Van Cor-Hosmer, S. K., et al. (2013). "Restricted 5'-end gap repair of HIV-1 integration due to limited cellular dNTP concentrations in human primary macrophages." *J Biol Chem* **288**(46): 33253-33262.

Viguera, E., et al. (2001). "Replication slippage involves DNA polymerase pausing and dissociation." *EMBO J* **20**(10): 2587-2595.

Wechsler, J. A. and J. D. Gross (1971). "Escherichia coli mutants temperature-sensitive for DNA synthesis." *Mol Gen Genet* **113**(3): 273-284.

Wijffelman, C. and P. van de Putte (1974). "Transcription of bacteriophage mu. An analysis of the transcription pattern in the early phase of phage development." *Mol Gen Genet* **135**(4): 327-337.

Willetts, N. S., et al. (1969). "Genetic location of certain mutations conferring recombination deficiency in *Escherichia coli*." *J Bacteriol* **97**: 244-249.

Withers, H. L. and R. Bernander (1998). "Characterization of *dnaC2* and *dnaC28* mutants by flow cytometry." *J Bacteriol* **180**: 1624-1631.

Wu, Z. and G. Chaconas (1995). "A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage." *EMBO J* **14**(15): 3835-3843.

Yamamoto, N., et al. (2009). "Update on the Keio collection of Escherichia coli single-gene deletion mutants." *Mol Syst Biol* **5**: 335.

Yang, Y. X., et al. (2010). "Cell context-dependent involvement of ATR in early stages of retroviral replication." *Virology* **396**(2): 272-279.

Yoder, K. E. and F. D. Bushman (2000). "Repair of gaps in retroviral DNA integration intermediates." *J Virol* **74**(23): 11191-11200.

Yoder, K. E., et al. (2011). "The base excision repair pathway is required for efficient lentivirus integration." *plos one* **6**(3): e17862.

Zavitz, K. H. and K. J. Marians (1992). "ATPase-deficient mutants of the Escherichia coli DNA replication protein PriA are capable of catalyzing the assembly of active primosomes." *J Biol Chem* **267**(10): 6933-6940.

Zavitz, K. H. and K. J. Marians (1993). "Helicase-deficient cysteine to glycine substitution mutants of Escherichia coli replication protein PriA retain single-stranded DNA-dependent ATPase activity. Zn2+ stimulation of mutant PriA helicase and primosome assembly activities." *J Biol Chem* **268**(6): 4337-4346.

Zhang, J. Z. (2003). "Evolution by gene duplication: an update." *Trends Ecol Evol* **18**(6): 292-298.

Zieg, J. and S. R. Kushner (1977). "Analysis of genetic recombination between two partially deleted lactose operons of *Escherichia coli* K-12." *J Bacteriol* **131**: 123-132.